



RAPPORT LNR 5123-2005

# Statistisk vurdering av overvåkning av dioksiner i organismer

Kvantifisering av usikkerhet og vurdering  
av utsagnskraft – grunnlag for planlegging  
av overvåkningsprogram



Norsk institutt for vannforskning

# RAPPORT


Hovedkontor	Sørlandsavdelingen	Østlandsavdelingen	Vestlandsavdelingen	Midt-Norge
Postboks 173, Kjelsås	Televeien 3	Sandvikaveien 41	Nordnesboder 5	Postboks 1264, Pilseneret
0411 Oslo	4879 Gvinstad	2312 Ottestad	5005 Bergen	7462 Trondheim
Telefon (47) 22 18 51 00	Telefon (47) 37 29 50 55	Telefon (47) 62 57 64 00	Telefon (47) 55 30 22 50	Telefon (47) 73 87 10 34 / 44
Telefax (47) 22 18 52 00	Telefax (47) 37 04 45 13	Telefax (47) 62 57 66 53	Telefax (47) 55 30 22 51	Telefax (47) 73 87 10 10
Internett: www.niva.no				


Tittel	Løpentr. (for bestilling)		Dato	
	5123-2005		13-02-2006	
Statistisk vurdering av overvåkning av dioksiner i organismer Kvantifisering av usikkerhet og vurdering av utsagnskraft – grunnlag for planlegging av overvåkningsprogram	Prosjektnr.	Undernr.	Sider	Pris
	O-24275		110	
Forfatter(e)	Fagområde		Distribusjon	
Bjerkeng, Birger	Geografisk område		Trykket	
	Vest-Agder og		NIVA	
	Telmark			


Oppdrags giver(e)	Oppdragsreferanse
Fylkesmannen i Vest-Agder	Dag Petter Sødal, Jon Egil Vinje

Sammenfatting
Eksisterende data for dioksin i organismer i Kristiansandsfjorden og Grenland er gjennomgått for å vurdere hva en kan oppnå av statistisk utsagnskraft ved fremtidig overvåkning i Kristiansandsfjorden, spesielt når det gjelder å følge med på tidsutvikling i dioksinkonsentrasjoner i biota etter de store utslippspreduksjonen som har skjedd i perioden 1990-2000. Noen dioksinkomponenter og kombinasjonsmål som virker særlig egnet til trendovervåkning er identifisert, ut fra miljømessige og statistiske vurderinger.
Størrelsen på tilfældige avvik fra jevne trender for enkelprøver er estimert til 30-40 % relativt standardavvik for blåskjell og 40-50 % for torskelever, skrubbefilet og krabbesmør. Reduksjonstrater på 5 til 10 % pr. år anslås som realistisk i perioder etter store utslippspreduksjoner, men sannsynligvis lavere for skrubbeflyndre. Estimaten er i hovedsak basert på det største datamaterialet fra Grenland.
Utsagnskraft som funksjon av omfang og varighet av overvåkning er estimert for ulike typer statistiske tester på hver organismetype. Det kan forventes at reduksjon siden 1996 kan påvises etter tre års ny overvåkning på 3-4 stasjoner, kanskje allerede etter første år for blåskjell. For å påvise fortsatt reduksjon vil det være nødvendig med 6-10 års årlig overvåkning på 4 stasjoner og 2-3 år lenger med bare to stasjoner. Med 2 til 5 år mellom hver observasjon på 4 stasjoner i stedet for årlig overvåkning vil nødvendig tidsrom forlenges med 2 til 5 år. Ved klassifisering i forhold til grenseverdier vil det kreve 4-10 års overvåkning på 4 stasjoner å påvise at kriteriet er oppfylt dersom reelle konsentrasjoner er omkring 0,6 til 0,8 ganger klassifiseringsgrensen.

Fire norske emneord	Fire engelske emneord
1. Miljøgifter	1. Environmental toxics
2. Dioksin	2. Dioxin
3. Biota	3. Biota
4. Grenland/Kristiansandsfjorden	4. Grenland/Kristiansandsfjord

  
Birger Bjerkeng  
Prosjektleder

  
Dominique Durand  
Forskningsleder

  
Øyvind Sørensen  
Ansvarlig

O-24275

**Statistisk vurdering av overvåkning av dioksiner i  
organismer**

- Kvantifisering av usikkerhet  
og vurdering av utsagnskraft  
– grunnlag for planlegging av overvåkningsprogram

## Forord

Denne rapporten er laget på oppdrag fra Fylkesmannen i Vest-Agder, for midler fra Statens forurensningstilsyn. Hensikten var å gjennomgå eksisterende data for dioksin i organismer i Kristiansandsfjorden og vurdere hva en kan oppnå av statistisk utslagskraft ved fremtidig overvåkning i Kristiansandsfjorden, spesielt når det gjelder å følge med på tidsutvikling i dioksinkonsentrasjoner i biota. Arbeidet bygget også i stor grad på datamaterialet som er samlet inn fra Grenland gjennom overvåkning utført av NIVA.

Kontaktperson hos Fylkesmannen i Vest-Agder har vært Dag Petter Sødal.

Oslo, 13. februar 2006

*Birger Bjerkeng*

# Innhold

<b>Sammendrag</b>	<b>6</b>
<b>1 Bakgrunn</b>	<b>16</b>
<b>2 Problemstilling</b>	<b>18</b>
<b>3 Beskrivelse av datamateriale</b>	<b>19</b>
3.1 Uivalg av datamateriale – geografiske områder og vevstyper	19
3.2 Geografisk og tidsmessig dekning i valgte områder	19
3.2.1 Grenland	20
3.2.2 Kristiansandsfjorden	20
3.3 Data om andre miljøgifter som støttevariable	22
3.4 Biologisk karakteristikk av prøvene	23
3.5 Avvikende verdier – utlukning av mulige feil.	24
3.6 Detekterbarhet for dioksinkomponenter – vurdering av hvilke enkeltkomponenter og kombinasjoner som egner seg best for statistisk analyse.	24
<b>4 Metode</b>	<b>27</b>
4.1 Aktuelle statistiske tester på nytt overvåkningsmateriale	27
4.2 Estimering av varianskomponenter	28
4.3 Bruke resultatene til å vurdere utsagnskraft	32
<b>5 Statistisk analyse av data for blåskjell</b>	<b>33</b>
5.1 Data fra Grenland	33
5.1.1 Beskrivelse av data	33
5.1.2 Statistiske analyser	35
5.2 Data fra Kristiansandsfjorden	37
5.2.1 Beskrivelse av data	37
5.2.2 Statistiske analyser	38
5.3 Sammenfattende konklusjon for blåskjelldata	41
<b>6 Statistisk analyse av data for torsk</b>	<b>43</b>
6.1 Data fra Grenland	43
6.1.1 Analyse av replikate prøver fra samme område og år	44
6.1.2 Analyse av stasjoner med lengre tidsserier	45
6.1.3 Sammenhenger med biologisk karakteristikk eller andre miljøgifter	48
6.2 Data fra Kristiansandsfjorden	50
6.2.1 Beskrivelse av data	50
6.2.2 Statistiske analyser	51
6.3 Sammenfattende konklusjon for torskelever	52
<b>7 Statistisk analyse av data for flét av skrubbeflyndre</b>	<b>54</b>
7.1 Grenland	54
7.1.1 Beskrivelse av data	54
7.1.2 Statistiske analyser	55
7.1.3 Sammenheng med andre miljøgifter	56

---

7.2	Kristiansandsfjorden	56
7.2.1	Beskrivelse av data	56
7.2.2	Statistiske analyser	57
7.3	Sammenfattende konklusjon for skrubbeflyndre	58
<b>8</b>	<b>Statistisk analyse av data for krabbe</b>	<b>59</b>
8.1	Data fra Grenland	59
8.1.1	Beskrivelse av data	59
8.1.2	Statistiske analyser	60
8.2	Data fra Kristiansandsfjorden	64
8.2.1	Beskrivelse av data	64
8.2.2	Statistiske analyser	65
8.3	Sammenfattende konklusjon for krabbe	67
<b>9</b>	<b>Utsagnskraft ved hypotesetesting som funksjon av omfang og design av overvåkingsprogram</b>	<b>69</b>
9.1	Varianskomponenter fra eksisterende data – en oppsummering	69
9.2	Vurdering av utsagnskraft ved statistiske tester av data fra ny overvåking	71
9.2.1	Utsagnskraft for tester på om dioksin-nivåene i organismer redusert siden 1996	72
9.2.2	Utsagnskraft for tester på om konsentrasjonene fortsatt reduseres	81
9.2.3	Utsagnskraft for å teste om gjennomsnittskonsentrasjoner ligger innenfor gitte klassifiseringsgrenser.	87
<b>10</b>	<b>Litteratur</b>	<b>91</b>
Vedlegg A.	Logtransformasjon	93
Vedlegg B.	Test-scenario – kovariansmodell	94
Vedlegg C.	Hypotesetesting – signifikansnivå og teststyrke	97
Vedlegg D.	Teststyrke for t-tester	98
Vedlegg E.	Utsagnskraft ved hypotesetesting av tidstrend og gjennomsnittsnivåer	99
Vedlegg F.	Beregning av varianskomponenter fra variansanalyser	104
Vedlegg G.	Presisjon for variansestimater som funksjon av antall frihetsgrader	106
Vedlegg H.	Statistiske egenskaper for blandprøver når konsentrasjoner er log-normalt fordelt.	107
Vedlegg I.	Definisjon av statistiske begreper	109

# Sammen drag

## Bakgrunn

I løpet av perioden 1989-1993 ble utslipp av bl.a. dioksin til indre deler av Kristiansandsfjorden fra Falconbridge Nikkelverk AS sterkt redusert fra omkring 3-6 g pr. år i 1988-89 til omkring 0.3 g/år fra og med 1993. Etter installering av elektrofilter i 2001 skal utslippene nå være nede på omkring 0.03 g/år, etter noe fluktuasjoner pga. driftsproblemer.

Analys er av biota i tidsrommet 1982-1996 har vist dioksin-nivåer som tilsier kostholdsråd og omsetningsrestriksjoner på sjømat fanget i området, selv om det ble påvist en nedgang i dioksin-konsentrasjonene i fisk ved siste prøvetaking i 1996. Et par blåskjellprøver fra Kristiansandsfjorden i 2002 samt erfaringen fra Grenlandsområdet etter store utslippsreduksjoner indikerer at konsentrasjonene i torskelever nå kan være kommet ned mot referansekonsentrasjonen, mens det for noen andre prøvetypene antagelig fortsatt vil forekomme høyere konsentrasjoner.

Det gjennomføres nå tiltak for å redusere sekundær belastning fra forurensede sedimenter, bl.a. ved overdekning. Det er da viktig med videre overvåkning for å se hvilken virkning tiltakene har hatt, revurdere kostholdsråd, og gi prognoser for fremtidig utvikling. I denne rapporten brukes eksisterende data for dioksin i biota fra Kristiansands-fjorden og Grenland til å estimere statistisk usikkerhet ved prøvetaking for å bestemme gjennomsnittskonsentrasjoner og endring over tid. På basis av dette vurderes hvordan et overvåkningsprogram bør gjennomføres for å oppnå rimelig utsagnskraft når det gjelder å påvise endringer.

## Problemstilling

Hvis det skjer en langsiktig endring i eksponering for miljøgifter over tid, vil nivåene i organismer i gjennomsnitt reflektere dette, men målte nivåer kan også variere ganske mye av andre årsaker, mer eller mindre uregelmessig ("tilfeldig") og på ulike rom- og tidsskalaer pga. reell variasjon i ytre miljøforhold eller i populasjonsdynamikk. I tillegg kommer utvalgsusikkerhet pga. variasjon mellom individer på samme tid og sted og evt. endringer over tid i prøvetakingsopplegg.

De uregelmessige variasjonene rundt en langsiktig jevn tidsutvikling (trend) eller et gjennomsnittlig nivå som er konstant i tid utgjør en "støy" i dataene. Støyen kvantifiseres ved *variansen*, som er middelværdien av kvadratet av disse variasjonene eller avvikene fra trend eller gjennomsnitt. *Standardavviket*, som er det vanlige målet for et 'typisk' avvik, er kvadratroten av variansen. Støyen gjør det vanskeligere å oppdage og kvantifisere langsiktige endringer over tid, fordi den gir tilfeldige utslag i middelværdier og koefisienter for tidsvariasjon. Ved vurdering av forskjeller mellom geografiske områder kan en justere for variasjoner i tid som er felles for områdene, mens uavhengige variasjoner i hvert område vil gi "støy" i beregnede forskjeller.

Ved statistisk testing sammenlignes beregnet effekt med det som kan forventes som restvirkning av de tilfeldige avvikene, og det vurderes om det er en reell effekt ut fra en aveiing av risikoen for å ta feil i den ene eller andre retningen. Hvis prøvetakingen kan gjøres på en slik måte at de tilfeldige variasjonene er små eller ujevnes effektivt, vil testen blir mer følsom for små effekter. Det oppnås først og fremst ved å utvide prøvematerialet, men også ved å velge et optimalt sett av målevariable, korrigere måleverdiene for ulike faktorer hvor det er en kjent sammenheng (fysiologi, årstid), velge prøvemateriale med lite tilfeldig variasjon i forhold til respons på ytre eksponering av miljøgifter, og ved å fordele prøvetakingen over tid og sted innenfor et år slik at de uregelmessige variasjonene fra år til blir jevnet ut best mulig.

Det er tilfeldig variasjon både mellom individer fra samme tid og sted og i gjennomsnittsnivåer for ulike prøvetakinger, dvs. forskjellige tidspunkt og/eller steder. Den siste typen variasjon kan være uavhengig fra sted til sted ved samme tidspunkt, eller med samvariasjon mellom steder.

Ved å bruke eksisterende data til å estimere hvor store de ulike bidragene til variasjon er og analysere hvordan de bidrar til variasjon i middelværdir kan en si noe om hvordan utsagnskraften varierer med overvåkingsprogrammets omfang, hyppighet og varighet. Dette kan gjøre det mulig å trekke konklusjoner om hva som er et optimalt overvåkingsprogram ut fra hvor små endringer en er interessert i å påvise i gitt tidsperspektiv og ut fra kostnader.

I denne rapporten er det gjort en slik analyse med sikte på å vurdere utsagnskraft ved fremtidig overvåkning av dioksiner i organismer i Kristiansandsfjorden.

## Datamateriale

Analysen baserer seg på data fra tidligere overvåkning i Kristiansandsfjorden og Grenlandsfjordene av blåskjell, torskelever, skrubbefilet og krabbesmør, stort sett ved blandprøver. Urvalgt materiale omfatter i alt 63 prøver fra Kristiansandsfjorden og 246 prøver fra Grenland. Prøvene er stedfestet både ved delområde og stasjonsnavn og tidfestet med år og måned. Noe data for biologisk karakteristikk av prøvene og andre klororganiske stoffer enn dioksin er tatt med som støttevariable.

Data er gjennomgått og noen få sterkt avvikende verdier er utelukket. Det er vurdert hvilke dioksin-komponentene som er mest egnet til trendanalyser ved at er godt målbare (ingen eller svært få prøver med verdier under analytisk kvantifiseringsgrense), samtidig som de har betydelige bidrag til toksisitetsekvivalent (TE). De mest egnede komponentene er funnet å være 2378-TCDF, 12378/12348-PeCDF, 23478-PeCDF, og analysen er konsentrert om kombinasjoner av disse tre (Sum bidrag til TE, aritmetisk eller geometrisk middelværdi). Ved senere statistiske analyser av overvåkningsresultater kan det selvsagt være aktuelt å utnytte mer informasjon.

## Aktuelle statistiske tester

Tidligere prøvetaking i Kristiansandsfjorden viste konsentrasjoner over SFTs klassifiseringsgrense for klasse I, men en nedgang over tid. Det er det naturlig å sette opp tre aktuelle spørsmål en vil ønske å få svar på ved ny prøvetaking:

1. Har det vært ytterligere nedgang i konsentrasjonene siden 1996, og hvor stor har nedgangen vært?
2. Endres konsentrasjonene fortsatt over tid, og eventuelt hvor raskt?
3. Hva er konsentrasjonsnivået i forhold til SFTs klassifiseringsgrense?

For alle de tre spørsmålene vil det være naturlig å se på nye og gamle data i sammenheng, og bruke resultatene til å estimere nå-situasjonen og evt. fortsette fremtidig utvikling ved fremskriving av observerte tidsserie. For de to siste spørsmålene kan det også være aktuelt å bare se på resultatene fra en ny overvåkning og vurdere resultatene uavhengig av eldre data.

For hvert av spørsmålene må utsagnskraft vurderes ved å anslå hvor store endringene må være for at de skal påvises med en rimelig høy sannsynlighet ved ulike omfang av ny overvåkning. Vurderingen er basert på estimert varians for ulike typer tilfeldige variasjoner ut fra eksisterende data.

## Hovedkonklusjoner

De dioksin-komponenter som er best kvantifisert og derfor mest egnet til statistisk analyse ser ut til å være 2378-TCDF, 12378/12348-PeCDF og 23478-PeCDF. I denne rapporten er det stort sett disse tre, eller kombinasjoner av dem, som er brukt til å estimere varianskomponenter og vurdere teststyrke for fremtidig overvåkingsprogram. Ved senere statistiske analyser av overvåkningsresultater kan det selvsagt være aktuelt å utnytte mer informasjon.

Datamaterialet fra Kristiansandsfjorden er sparsomt, med prøvetaking bare for noen få tidspunkter over en periode med store utslippsreduksjoner. Det har derfor vært nødvendig å bygge på det mye



større materialet fra Grenland for å estimere varianskomponenter og tidstrender. De estimatene det er mulig å gjøre for data fra Kristiansandsfjorden avviker ikke signifikant fra de mer presise estimatene fra Grenlandsfjordene, og det antas derfor at de siste også kan brukes for Kristiansandsfjorden.

Dioksin-konsentrasjoner i blandprøver av typisk størrelse fra forskjellige organismetyper (50 blåskjell, 20 torskelever, skrubbefilet eller krabbesmør) har omtrent samme grad av tilfeldige avvik fra år til år rundt jevne tidstrender. Avvikene tilsvarer standardavvik på ca. 35 til 55 % for enkeltkonsentrasjon (Variansen omkring 0.18 på naturlig logskala). Prøver fra forskjellige stasjoner ved samme tidspunkt har avvik som i stor grad er uavhengig av hverandre, og de vil derfor jevne seg ut i gjennomsnitt beregnet over flere stasjoner.

Det er ikke funnet grunnlag for å fastslå noen påliteelig sammenheng med biologisk karakteristikk for blåskjell, torskelever eller skrubbefilet. For krabbesmør er det ganske klar sammenheng med kjønn og fett-%, og særlig korreksjon for kjønn har betydning. Forsøk på å sette dioksinkonsentrasjonene i sammenheng med andre klororganiske miljøgifter for eventuelt å kunne utnytte det til å forbedre utsagnskraften i overvåkingen har ikke gitt resultater.

Det har blitt vurdert om en kan øke utsagnskraft pr. kostnad ved å ikke samle alt materiale til en blandprøve ved samme sted og tid, men i stedet lage blandprøver av materiale fra flere steder rundt en stasjon, eller ved flere tidspunkter innenfor et år. På den åten kan en jevne ut variasjon i liten tid- og rom-skala og redusere tilfeldige avvik i data, men med de økte kostnader det medfører i følge regne-eksemplene i denne rapporten, ser det ut til å gi liten gevinst. Det kan evt. vurderes nærmere for et konkret overvåkningsopplegg hvor kostnadselementene er bestemt bedre.

**For å påvise nedgang siden 1996** i Kristiansandsfjorden ser det ut til at det kan være nok med overvåking over 3 år for de fleste prøvetypene, hvis det har vært en reduksjon på 8-10 % pr. år. Ut fra data fra Grenland virker en slik reduksjonsrate realistisk for en 10-årsperiode etter de store utslipps-reduksjoner som har foregått også i Kristiansandsfjorden. Det er da forutsatt prøvetaking på omtrent de samme stedene som tidligere, og med samme prøvestørrelse. Det kan være mulig å påvise endringer allerede etter første år, spesielt for blåskjell, men tre år er et minimum for å estimere størrelsen på uregelmessige fluktusjoner i dagens situasjon. For blåskjell bør det tas prøver på både Myrrodden og Fiskåtangen i område A, og både ved Silokaia og Odderøy vest i område B. For de andre prøvetypene bør det tas prøver av område B, D og E.

**For å se om det fortsatt pågår en reduksjon** vil det være aktuelt å se bare på data fra ny overvåking. Da må en regne med et lengre overvåkningsprogram. Hvis en er interessert i å påvise reduksjon med stor sannsynlighet hvis faktisk reduksjonsrate er 5-10 % pr. år, kan det kreves årlig overvåking på 8-10 år på 4 stasjoner, eller kanskje ett til to år kortere hvis det er uavhengige fluktusjoner fra år til år på ulike stasjoner. Med bare 2 stasjoner må en regne med 2-3 års lenger overvåkningsperiode før en kan påvise slike reduksjonsrater. Ved å øke antall stasjoner kan en påvise trender for noe kortere tidsrom, men til høyere total kostnad, og forbedringen kan begrenses av samvariasjon mellom mer tettliggende stasjoner. For Kristiansandsfjorden kan 3-4 stasjoner være et optimalt antall med hensyn til en slik vurdering.

Jo kortere tidsrommet er, jo mindre er sjansen for skille fra hverandre midlertidige svingninger av noen års varighet og langsiktig trend. En overvåkningsperiode på 8-10 år kan anses som et minimum for å vurdere tidstrender.

Med prøvetaking hvert annet, tredje eller femte år i stedet for hvert år forlenges nødvendig tidsrom for å påvise trend fra 10 år til hhv. 12, 13 og 15-20 år ved overvåking på 4 stasjoner. Samtidig kan totale kostnader bli 30-60 % mindre ut fra de kostnadselementene som er brukt her, men dette må vurderes nærmere for et konkret overvåkningsopplegg.

Hvis en setter større krav til nøyaktig estimat for tidstrenden ut over det som skal til for å si at det er en reduksjon, øker kravet til overvåkningsperiode. Ved årlig overvåking på 4 stasjoner kreves en periode på rundt 15 år før en kan regne med å estimere reduksjonsraten innenfor et 90 % konfidens-

intervall på  $\pm 2.0$  % pr. år. Tilsvarende lengre periode kreves også for bare å påvise at det skjer en reduksjon hvis den reelle reduksjonsraten skulle være lavere enn 5-10 % pr. år.

**For at en skal kunne påvise statistisk at gjennomsnittskonsentrasjoner ligger innenfor gitte klassifiseringsgrenser** må reell gjennomsnittskonsentrasjon være mindre enn grenseverdien, og forskjellen må være større jo mindre data en har, dvs. jo større usikkerhet det er i estimert konsentrasjon.

Med en enkel serie av årsmidler basert på data over 4 år fra 4 stasjoner må forholdstallet<sup>1</sup> mellom reell gjennomsnittskonsentrasjon og klassifiseringsgrensen være mindre enn 0.66 for at en skal ha teststyrke 90 % for å konkludere med at konsentrasjonene er innenfor grensen ved ensidig test med signifikansnivå 10 %. Hvis det påvises signifikant tidstrend bør klassifiseringen gjøres ved å sammenligne grenseverdien med estimert gjennomsnitt ved slutten av 4-årsperioden i følge denne tidstrenden. Da må reell gjennomsnittskonsentrasjon ved slutten av perioden være under halvparten av klassifiseringsgrensen for at teststyrken skal være den samme.

Etter 10 års overvåking hvert år på 4 stasjoner oppnås samme utsagnskraft for forholdstall opp mot 0.8 uten tidstrend og 0.65 ved slutten av perioden med tidstrend. Med data over 10 år fra bare to stasjoner kreves 6 års lenger overvåking for å oppnå samme evne til å klassifisere, men forskjellen i krav til forholdstallet mellom 10 og 16 års overvåking er ikke så stor.

Hvis stasjonene har stor grad av uavhengige avvik fra gjennomsnitt eller trend, slik at data kan analyseres som uavhengige observasjoner fra ulike kombinasjoner av stasjon og år, vil evnen til å klassifisere riktig forbedres. Det gjelder spesielt for de kombinasjonene som gir krav til lavt forholds-tall hvis data analyseres på en enkel tidsserie aggregert over stasjoner. Med data over 4 år fra 4 stasjoner vil utsagnskraft som beskrevet oppnås hvis konsentrasjon uten tidstrend er mindre enn 0.77 ganger klassifiseringsgrensen, og med 10 års overvåking opp mot 0.85. Hvis det er signifikant tidstrend, må konsentrasjonen ved slutten av perioden være innenfor hhv. 0.72 og 0.81 ganger klassifiseringsgrensen.

## Mer detaljert gjennomgang av resultatene

### Estimering av varianskomponenter

Varianskomponenter er estimert separat for hver prøvetype ved ulike former for varians- og regresjonsanalyse. Alle analyser er gjort på naturlige logaritmer av konsentrasjoner. Med en slik transformasjon av data kan forskjeller i ytre miljø og biologisk betingede faktorer skilles fra hverandre i additive statistiske modeller (med en sum av ulike ledd) selv om faktorene virker multiplikativt (f.eks. hvis konsentrasjon i organismer = konsentrasjon i ytre miljø multiplisert med akkumulerings-faktor). Log-transformering gir også ofte fordelinger som er mer symmetriske, og hvor variansen er mer uavhengig av gjennomsnittlig nivå, slik de vanlige statistiske modellene krever.

De statistiske analysene av tidstrender for dioksiminnhold basert på datamaterialet fra Kristiansandsfjorden og Grenlandsfjordene gir følgende resultat:

### Blåskjell:

- Hvis tidsutviklingen i dioksinkonsentrasjoner på ett sted overvåkes ved at en tar blåskjell-prøver over en årrekke med 50 skjell i hver prøve, vil konsentrasjoner for enkeltprøvene variere tilfeldig fra år til år rundt en jevn tidstrend. Disse variasjonene har relativt standard-avvik ca. 30 til 40 % (variens 0.11 for naturlige logaritmer av konsentrasjoner).

---

<sup>1</sup> Basert på testing av logtransformerte konsentrasjoner, men regnet tilbake til lineære konsentrasjoner.

- Omtrent halyparten av variansen ser ut til å være utvalgsusikkerhet pga. tilfeldig variasjon mellom skjell fra samme tid og sted. Den andre halyparten skyldes antagelig reelle avvik fra trenden i gjennomsnittet fra år til år. Hvis det analyseres flere prøver fra samme stasjon hvert år, eller hvis antall individer pr. prøve økes, vil virkningen av utvalgsusikkerheten avta, mens virkningen av de reelle avvikene vil være den samme.

- Hvis det tas parallelle tidsserier fra flere stasjoner innenfor et område, kan en regne med at disse avvikene fra tidstrendene er uavhengige på hver stasjon. Virkningen på samlet års-gjennomsnitt vil da minke jo flere stasjoner en har. I tillegg kan det være parallelle avvik over hele området som er felles for alle stasjoner (varians 0.025 på naturlig logskala). Dette variansbidraget vil ikke ha særlig betydning totalt sett hvis antall stasjoner er 6 eller mindre.
- Data fra Grenland viser ca. 10 % reduksjon av blåskjellkonsentrasjoner pr. år over en 10-års periode etter store utslippsreduksjoner. Det antas at det er realistisk å vente det samme i Kristiansandsfjorden etter de store utslippsreduksjonene som har skjedd der siden 1990.

#### **Torskelever:**

- Hvis tidsutviklingen i dioksinkonsentrasjoner på ett sted overvåkes ved at en tar prøver av torskelever over en årrekke med 20 fisk i hver prøve, vil konsentrasjoner for enkelprøver variere tilfeldig fra år til år rundt jevn tidstrend eller langsiktig uendret gjennomsnitt. Variasjonene har relativt standardavvik ca. 40 til 50 % (varians 0.16 for naturlige logaritmer av konsentrasjon).
- Noe over halyparten av variansen ser ut til å være utvalgsusikkerhet pga. tilfeldig variasjon mellom individuelle torsk fra samme sted og tid (varians 0.09 for naturlige logaritmer av konsentrasjon). Resten skyldes antagelig reelle avvik fra trenden i gjennomsnittet fra år til år (varians 0.07 for naturlige logaritmer av konsentrasjon). Hvis det analyseres flere prøver fra samme stasjon hvert år, eller hvis antall individer pr. prøve økes, vil virkningen av utvalgs-usikkerheten avta, mens virkningen av de reelle avvikene vil være den samme.
- Hvis det tas parallelle tidsserier fra flere stasjoner innenfor et område, kan en regne med at disse avvikene fra tidstrendene er uavhengige på hver stasjon. Usikkerheten i samlet års-gjennomsnitt vil da minke jo flere stasjoner en har. I tillegg kan det være avvik fra tidstrend som er felles for alle stasjoner innenfor et større område (estimert varians 0.02 på naturlig logskala). Dette variansbidraget vil ikke ha særlig betydning totalt sett hvis antall stasjoner er 4 eller mindre.
- Data fra Grenland viser ca. 8 % reduksjon av konsentrasjoner i torskelever pr. år over en 10-års periode etter store utslippsreduksjoner. Det antas at det er realistisk å vente det samme i Kristiansandsfjorden etter de store utslippsreduksjonene som har skjedd der siden 1990.
- Det er ikke funnet sikre sammenhenger med biologisk karakteristikkk eller andre klororganiske miljøgifter, og analyse av utsagnskraft bør ikke baseres på at det er mulig å utnytte dette.

#### **Skrubbeflyndre:**

- Hvis tidsutviklingen i dioksinkonsentrasjoner på ett sted overvåkes ved at en tar prøver av filét fra skrubbeflyndre over en årrekke med 20 fisk i hver prøve, vil konsentrasjoner for enkeltprøver variere tilfeldig fra år til år rundt jevn tidstrend. Variasjonene tilsvarer et relativt standardavvik på ca. 40-50 % for enkeltkonsentrasjoner (varians 0.15 for naturlige logaritmer av konsentrasjoner).
- Avvikene skyldes delvis utvalgsusikkerhet pga. tilfeldig variasjon mellom enkeltfisk fra samme sted og tid, men noe kan også skyldes reelle avvik i gjennomsnittet fra år til år i

forhold til langsiktig trend. Datagrunnlaget for skrubbefilét er ikke stort nok til at en kan tallfeste betydningen av utvalgsusikkerhet i forhold til reelle avvik.

- Hvis det tas prøver fra flere delområder hvert år, kan avvikene antas å være uavhengige for hvert delområde. Usikkerheten for samlet årsgjennomsnitt vil da minske jo flere delområder en har prøver fra. Det er ikke mulig å påvise avvik mellom år som er felles over flere stasjoner for det eksisterende datamaterialet.
- For gitt tid og sted ser konsentrasjonene ut til å variere systematisk omtrent proporsjonalt med fett-innholdet. Standardavviket for tilfældige avvik som er angitt ovenfor gjelder etter at det er korrigert for variasjon i fett-innholdet mellom prøvene.
- I Grenlandsområdet kan det ha vært en reduksjonsrate på opp til ca. 6 % pr. år, men mest sannsynlig mindre enn dette, og usikkerheten er for stor til at en kan påstå at konsentrasjonene i skrubbeflyndre er redusert i det hele tatt i perioden fra 1990 til 2000.

#### Krabbesmør:

- Hvis tidsutviklingen i dioksinkonsentrasjoner på ett sted overvåkes ved at en tar blandprøver av krabbesmør med ca. 20 individer pr. prøve over en årrekke, vil konsentrasjoner for enkeltprøvene variere tilfeldig fra år til år rundt jevn tidstrend. Disse variasjonene har relativt standardavvik 40 til 50 % (varians 0.17 for naturlige logaritmer av konsentrasjoner).
- Opp mot halvparten av variansen ser ut til å være utvalgsusikkerhet pga. tilfeldig variasjon mellom individer fra samme sted og tid (varians 0.08 for naturlige logaritmer av konsentrasjon). Resten skyldes antagelig reelle avvik fra tendens i gjennomsnittet fra år til år (varians 0.09 for naturlige logaritmer av konsentrasjon). Hvis det analyseres flere prøver fra samme stasjon hvert år, eller hvis antall individer pr. prøve økes, vil virkningen av utvalgsusikkerheten avta, mens virkningen av de reelle avvikene vil være den samme.
- Hvis det tas parallelle tidsserier fra flere stasjoner innenfor et område, kan en regne med at disse avvikene fra tidstrendene er uavhengige på hver stasjon. Virkningen på samlet årsgjennomsnitt vil da minske jo flere stasjoner en har. I tillegg ser det ut til å være avvik fra tidstrend som er felles for alle stasjoner innenfor et område (estimert varians 0.05 på naturlig logskala). Et slikt variansbidrag får betydning hvis det samles data på mer enn 2 stasjoner.
- Den variasjonen som er angitt ovenfor gjelder for konsentrasjoner korrigert for kjønn og fett-%. Kjønn er den viktigste faktoren å korrigere for. Det er viktig å kontrollere og registrere dette ved prøvetaking.
- Data fra Grenland viser ca. 9 % reduksjon av konsentrasjoner i krabbesmør pr. år over en 10-års periode etter store utslippsreduksjoner. Det antas at det er realistisk å vente det samme i Kristiansandsfjorden etter de store utslippsreduksjonene som har skjedd der siden 1990.

Resultatene for de forskjellige prøvetypene er svært like når usikkerheten i estimatene tas i betraktning. Forholdstall mellom variansestimater for forskjellige prøvetyper varierer omtrent som en vil vente for forskjellige estimater for samme varians. Det er derfor ikke noe sikkert grunnlag for å peke ut noen av prøvetypene som mer egnet enn andre til å detektere trender i dioksinkonsentrasjonene. Resultatene kan likevel peke i retning av at blåskjellprøvene har noe mindre fluktusjoner enn de andre prøvetypene, men det er marginalt i forhold til usikkerheten i estimatene.

Når formålet er å sammenligne med tidligere data er det dessuten viktig å kunne bruke så mye som mulig av det tidligere innsamlende materialet, spesielt siden det finnes så lite data, så det kan være en grunn til å ta nye prøver av alle de prøvetypene som det er tidligere data for i noe omfang.

Oppdelingen i ulike varianskomponenter er usikker, men de uavhengige analysene på ulike prøvetyper gir nokså like resultater, så det kan være grunn til å tro at oppdelingen er brukbar for å skissere utsagnskraft. Analyse av utsagnskraft og overvåkningsprogram generelt for alle prøvetyper baseres på et gjennomsnitt av de resultatene for de forskjellige prøvetypene.

### Vurdering av utsagnskraft ved statistiske tester av data fra ny overvåkning

Variansestimatene er mest basert på data fra Grenlandsområdet, men stemmer med mer usikre estimater fra Kristiansandsfjorden, så det er antatt at estimatene fra Grenland kan brukes for Kristiansandsfjorden. Varians mellom replikate prøver tatt på samme tid og sted kan ventes å være generelt anvendbar, mens variasjon mellom år og mellom kombinasjoner av år og stasjon kan være mer avhengig av konkrete forhold i hvert område. Varianskomponentene gir uansett bare en forenklet beskrivelse av en sammensatt variasjon på mange rom- og tids-skalaer. Det er ikke grunnlag for å vurdere dette nærmere.

Utsagnskraft som funksjon av omfang av nytt overvåkningsprogram er vurdert for tre ulike typer statistiske tester, tenkt anvendt på data for hver prøvetype separat:

#### Utsagnskraft for å teste om dioksinkonsentrasjoner i organismer er redusert siden 1996

Et viktig mål med ny overvåking i Kristiansandsfjorden vil være å se om det har skjedd en nedgang i nivåene siden 1996 vurdert ut fra nye dioksimålinger i organismer i 2006-2007. Spørsmålet er hvor stor en eventuell ending fra 1996-nivåene må være for at den skal kunne påvises noenlunde sikkert (med 80-90 % sannsynlighet) ved ny overvåking i 1-3 år. Vurderingen er basert på en ensidig test av om konsentrasjoner er redusert. Signifikansnivået er satt til 10 %, dvs. at sannsynligheten for å trekke gal konklusjon hvis det ikke har vært noen reduksjon, holdes innenfor 10 %.

- Dersom konsentrasjonene i **blåskjell** har blitt redusert med 50 % eller mer siden 1996, vil ett års ny overvåkning på 3-4 stasjoner antagelig kunne påvises reduksjonen rimelig sikkert. Hvis ny overvåkning forlenges til to eller tre år, kan en regne med å påvise også en nedgang på 40-45 % med samme sikkerhet. Hvis det er fluktasjoner fra år til år som er felles over stasjoner, kan det redusere evnen til å detektere forskjeller. Dette må vurderes når nye resultater foreligger.  
Siden reduksjoner på 65 % i løpet av 10 år kan være realistisk (10 % reduksjon pr. år), er det mulig at en vil kunne påvise signifikant ending allerede etter ett års ny overvåkning av blåskjell, forutsatt at en tar prøver fra 3 eller 4 stasjoner.  
For å få et bilde av variansen i dagens situasjon bør det tas sikte på iallfall tre års overvåkning, men det er altså mulig at en kan trekke foreløpige konklusjoner allerede etter det første året.
- Etter ett års ny overvåkning av **torskelever** med prøvetaking fra område B og D (Vesterhavna og Vestergapet), kan en regne med nokså sikkert å påvise en reduksjon fra 1996 dersom konsentrasjonene er redusert med minst 60-70 %. Fortsetter overvåkingen i ett eller to år til, vil det være mulig å påvise en reduksjon på 45-55 %.  
En reduksjonsrate ca. 8 % pr. år som er beregnet for torskelever i Grenland vil bety en nedgang med ca. 55 % i løpet av 10 år. Hvis det har vært en slik reduksjonsrate også i Kristiansandsfjorden vil det derfor kreve to, og helst tre års overvåkning å fastslå rimelig sikkert at det har vært en nedgang.  
Det er også uansett nødvendig med tre år for å få et inntrykk av hvor stabile nivåene er fra år til år og se om variansanalysen må ta hensyn til variasjon fra år til år som gjennomsnitt over områder. Det anbefales å ta prøver også fra område E (Dvergøy-Havøy), fordi dette kan komme til å styrke muligheten til å konkludere, men det er ikke tatt i betraktning her, fordi det er indikasjoner på at tidsutviklingen i område E er forskjellig fra B og D.

- For å påvise endring etter ett års overvåking av **skrubbefilet** fra område B og D med enveis test for reduksjon som skissert må konsentrasjonene ha vært redusert med anslagsvis 80-90 % siden 1996. Først etter tre års overvåking kan en regne med på å kunne påvise endring dersom reduksjonen har vært omkring 50-60 %. For di det uansett vil være snakk om svært lite data fra 1996 som sammenligningsgrunnlag, er disse vurderingene spesielt usikre. Estimert reduksjonstrater er enten ikke signifikant forskjellig fra null (Grenland år 1991-2000) eller svært usikre (Kristiansandsfjorden år 1999-1996), og det er vanskelig å vurdere hvor rimelig det er å anta at en vil kunne påvise reduksjon i løpet av ett til to års ny overvåking.

- For at en skal kunne påvise reduksjon fra 1996 noenlunde sikkert med enveis testing etter ett års overvåking av **krabbesmør** i område B, D og E må konsentrasjonene være redusert med anslagsvis 60-70 % fra 1996 til 2006. Etter to års overvåking kan en regne med å påvise reduksjon på 50-55 % og etter tre år reduksjon på omkring 45-50 %.

I Grenland ser det ut til at reduksjonstraten er ca. 9 % pr. år, dvs. 60 % over 10 år, og det kan derfor være rimelig å anta at en vil kunne påvise endring etter to års overvåking i Kristiansandsfjorden.

Det er viktig at fordeling på hann- og hunnkrabbe i prøvene registreres nøyaktig i datamaterialet, helst ved at det analyseres separate blandprøver av begge kjønn. Fett-% bør måles som kovariat, selv om det ikke har så stor betydning som kjønn.

### Utsagnskraft for å teste om konsentrasjonene fortsatt reduseres

Evnen til å påvise endring over tid er beregnet for overvåkningsprogram med ulike kombinasjoner av antall overvåkingsår og tidsintervall mellom dem, antall stasjoner og replikater pr. kombinasjon av stasjon og år. Det er også tatt med hva en vinner ved å fordele innsamling til hver blandprøve på litt ulike steder og/eller tidspunkter innenfor stasjon og år. Vurderingene av utsagnskraft er gjort for ett sett av varianskomponenter, uten å skille mellom ulike prøvetyper. Det er også gjort noen kostnads-estimer for ulike overvåkningsprogrammer, og da presenteres separate tall for krabbe og torskelever. Kostnadsberegningene er bare eksempler på slike beregninger, mer reelle vurderinger må gjøres i sammenheng med konkret planlegging av et overvåkningsopplegg, hvor feltkostnadene kan spesifiseres.

En tidstrendanalyse basert på årlige prøver fra 4 stasjoner, hvor data slås sammen til en felles tidsserie før analyse av tidstrend, vil kreve overvåking i 8-10 år eller mer før en kan regne med å påvise noenlunde sikkert en reduksjonstrate på 5-10 % pr. år. En slik reduksjonstrate tilsvarer omtrent det en finner i tidligere data fra Grenland.

Analysen av data fra Grenland tyder på at avvik innenfor samme år i stor grad er uavhengige fra stasjon til stasjon. Hvis dette også gjelder for Kristiansandsfjorden, kan data analyseres som uavhengige observasjoner i en modell med felles tidstrend for stasjonene, justert for konstant forskjell mellom stasjoner. Hvis testene kan baseres på dette, vil en reduksjonstrate på 5-10 % pr. år, felles over alle stasjonene, kunne påvises ett til to år før. Analysen kan alltid gjøres på en enkel aggregert tidsserie av gjennomsnitt over flere stasjoner, men hvis avvikene er uavhengige fra stasjon til stasjon vil dette være noe konservativt i forhold til å påvise en felles tidstrend. Som en mellomting kan en estimere graden av samvariasjon mellom stasjoner og la det inngå i testen for trend.

Hvis antall stasjoner reduseres fra 4 til 3, vil det ta ett år mer å oppdage samme trend, og med bare 2 stasjoner vil det ta to eller tre år mer. Virkningen av å redusere antall stasjoner blir størst hvis det er uavhengige avvik fra stasjon til stasjon. Samtidig kan totalkostnadene frem til en kan påvise reduksjon ved 3 og 2 stasjoner bli hhv. 15 og 30 % lavere sammenlignet med 4 stasjoner.

Omvendt vil en doubling fra 4 til 8 stasjoner gi 1 til 2 års raskere påvisning av en slik trend, men samtidig øker total kostnad med 50 til 70 %. Også her er gevinsten størst hvis det er uavhengige avvik fra stasjon til stasjon. Det er uansett usikkert hva en vinner med så mange stasjoner over et begrenset område, siden de kan ha større samvariasjon jo tettere de ligger. For det området av Kristiansands-

fjorden som har vært overvåket hittil kan 3-4 stasjoner være et optimalt antall med hensyn til en slik vurdering.

Med prøvetaking hvert annet år i stedet for hvert år kreves et 2 år lengre tidsrom før en kan påvise en reduksjon på ca. 6 % pr. år like sikkert; med prøvetaking hvert tredje år kreves et 3 år lengre tidsrom. Hvis det tas prøver bare hvert femte år kan det ta 5 til 10 år mer før en kan påvise samme trend som for årlig overvåking etter ca. 10 år. Totale kostnader for å påvise samme reduksjonsrate blir 30-60 % mindre ved disse alternativene sammenlignet med årlig prøvetaking. Alle disse sammenligningene gjelder 4 stasjoner.

Dersom en kan finne 6 prøvetakingssteder med uavhengige variasjoner, og analyserer 3 blandprøver basert materiale fra to steder, kan det gi samme utsagnskraft som enkelprøver fra 4 stasjoner, og til ca. 10 % mindre kostnad over samme tidsrom ut fra de kostnadselementene som er satt opp som eksempler, men forskjellen må betraktes som usikker.

### **Utsagnskraft for å kunne påvise at gjennomsnittskonsentrasjon ligger innenfor gitte klassifiseringsgrenser.**

Det kan også være aktuelt å bruke overvåkingsdata til å klassifisere et område i forhold til definerte grenseverdier. For å kunne fastslå med rimelig sikkerhet at konsentrasjonen er under en gitt klassifiseringsgrense, må reell gjennomsnittskonsentrasjon være mindre enn grenseverdien, og forskjellen må være større jo mindre data en har, dvs. jo større usikkerhet det er i estimert konsentrasjon.

Hvis data over 4 år fra 4 stasjoner aggregeres til en middelerverdi pr. år, og det ikke er noen tidstrend, må gjennomsnittlig konsentrasjon være mindre enn 0.66 ganger klassifiseringsgrensen for at en ensidig test skal ha teststyrke 90 %, dvs. ha 90 % sannsynlighet for å konkludere med at klassifiseringskriteriet er oppfylt, når signifikansnivå settes til 10 %. Hvis konsentrasjonene endres over tid, og grensen sammenlignes med estimert gjennomsnitt ved slutten av 4-årsperioden, må reell konsentrasjon da være nede i under halvparten av klassifiseringsgrensen for å oppnå det samme.

Etter 10 års overvåking hvert år på 4 stasjoner oppnås samme utsagnskraft hvis reell konsentrasjon er mindre enn 0.8 ganger klassifiseringsgrensen dersom det ikke er noen tidstrend. Hvis det er en tidstrend, må reell konsentrasjon være under 0.65 ganger klassifiseringsgrensen ved slutten av perioden for å oppnå samme klassifiseringsikkerhet. Med data fra bare to stasjoner kreves 3 års lenger overvåking for å oppnå samme evne til å klassifisere.

Hvis stasjonene har stor grad av uavhengige avvik fra gjennomsnitt eller trend, slik at data kan analyseres som uavhengige observasjoner fra ulike kombinasjoner av stasjon og år, vil evne til å klassifisere riktig forbedres. Med data over 4 år fra 4 stasjoner vil utsagnskraft som beskrevet oppnås hvis konsentrasjon uten tidstrend er mindre enn 0.77 ganger klassifiseringsgrensen, og med 10 års overvåking opp mot 0.85. Hvis det er signifikant tidstrend, må konsentrasjonen ved slutten av perioden være innenfor hhv. 0.72 og 0.81 ganger klassifiseringsgrensen.

### **Forutsetninger og forbehold**

Anslått utsagnskraft mht. å påvise endringer over tid er basert på følgende forutsetninger:

- Det gjøres ensidig test av om det har vært reduksjon, dvs. nullhypotesen 'Gjennomsnittskonsentrasjoner er uendret eller øker med tiden' testes mot alternativ hypotese 'Gjennomsnittskonsentrasjoner reduseres med tiden'.
- Signifikansnivå settes til 0.1, dvs. det aksepteres en risiko på 10 % for å konkludere feilaktige med at det har vært reduksjon selv om det ikke er tilfelle.

- For å si at en endring kan påvises med stor sannsynlighet kreves at teststyrken for slike tester skal være 80-90 %.
- Måleverdier er omtrent lognormalt fordelt rundt forventningsverdier ut fra stasjonsgjennomsnitt og tidstrender.

Tilsvarende forutsetninger gjelder for vurdering av utsagnskraft ved klassifisering, men da med nullhypotese: 'gjennomsnittskonsentrasjoner er høyere enn klassifiseringsgrensen' mot alternativ hypotese 'gjennomsnittskonsentrasjoner er innenfor klassifiseringsgrensen'.

Hvis kravene til treffsikkerhet økes (lavere signifikansnivå, større teststyrke), eller en vil bruke tosidige tester hvor en også er interessert i å påvise eventuell økning over tid, vil kravet til overvåkningen omfang og varighet øke.

Det er antatt at avvikene fra tidstrend for prøver fra én stasjon er uavhengige fra år til år. Dersom avvikene i stor grad opptrer som mer langvarige svingninger, for eksempel av to-tre års varighet, vil det ta lenger tid enn angitt her å påvise langsiktige trender.

For tester på et datasett som omfatter flere stasjoner, og hvor stasjon inngår som faktor i analysen, er det i denne rapporten antatt at avvikene fra jevne tidstrender ved hvert tidspunkt er uavhengige fra stasjon til stasjon. Analysen av data fra Grenland viser at slik felles variasjon over flere stasjoner for blåskjell, torsk og skrubbe ikke er signifikant, og antagelig så liten at den ikke har særlig betydning så lenge antall stasjoner er 4 eller mindre viktig. For krabbesmør ser det ut til å være noe mer slik variasjon, og da vil gevinsten ved å øke antall stasjoner ut over 2 blir da mindre. For Kristiansandsfjorden er det for lite data til å gi noe bilde av dette, men en analyse på blåskjelldata derfra gir ikke signifikante felles avvik fra tidstrend over flere stasjoner. I den grad det er slik samvariasjon i avvik mellom stasjoner, vil ikke prøvene fra ulike stasjoner ved samme tidspunkt representere uavhengige målinger. En alternativ test, hvor data fra ulike stasjoner slås sammen til en enkel serie av årsmidler, representerer da en mer konservativ antagelse, hvor en ikke regner med at stasjonene fluktuere uavhengig av hverandre. Reimingslinjer basert på dette er det som best sikrer at overvåkningen faktisk når målene om utsagnskraft.

Når teststyrke beregnes, er det også antatt at tidstrend (som relativ reduksjon pr. år ved analyse på log-skala) er omtrent lik på alle stasjoner som analyseres sammen. Hvis ulike stasjoner har påvisbart forskjellige tidstrender, må en ta hensyn til det i analysen, og da vil teststyrken bli noe svakere, og det vil kreve noe lenger tid å estimere de ulike trendene pålitelig, enn det som er angitt i rapporten.

Vurdering av utsagnskraft vil alltid ha en viss usikkerhet, fordi estimater for varians har usikkerhet, og fordi en ikke kan vite sikkert om trender og avvik fra trender vil ha samme størrelse og karakter i en fremtidig utvikling. Forutsatt at det siste er oppfylt, kan det antas at størrelse på endringer eller forskjeller som kan påvises ikke avviker mer enn  $\pm 20$  % av det som er anslått.

De vurderingene som er gjort her er basert på at data for de ulike prøvetypene analyseres statistisk hver for seg, fordi det er antatt at den primære interessen er å følge med på utviklingen av konkrete organismetyper, bl.a. når det gjelder kostholdsråd. Hvis det tas prøver parallelt for flere organismetyper, kan en antagelig få bedre utsagnskraft for en konklusjon om generell reduksjon over tid ved å bruke alle data i en samlet analyse.

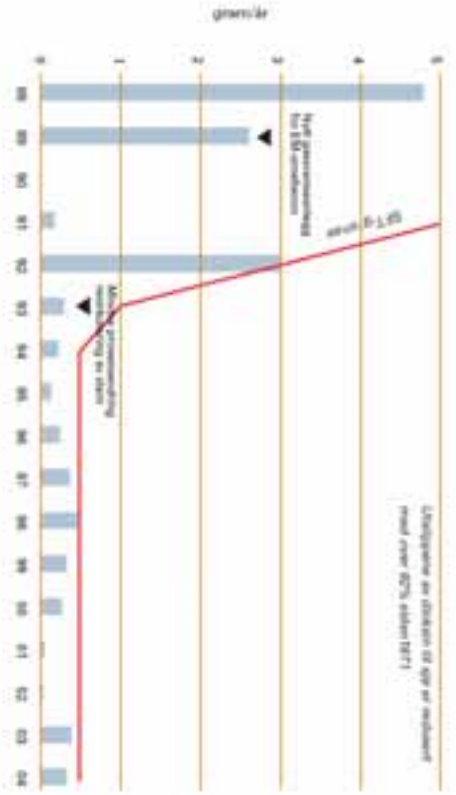


# 1 Bakgrunn

I løpet av perioden 1989-1993 ble utslipp av bl.a. dioksin til indre deler av Kristiansandsfjorden fra Falconbridge Nikkelverk AS sterkt redusert. Tabell 1 og **Figur 1** viser hvordan utslippene har utviklet seg ifølge tilgjengelige opplysninger. Det var en markert nedgang med en faktor 20 fra 1988-89 med utslipp på 3-6 g pr. år frem til perioden 1993-2000 da utslippene varierte omkring 0.3 g pr. år. Etter installering av elektroflter sank utslippene i 2001 og 2002 ytterligere med en faktor ca. 10 til hhv. 0.04 og 0.03 g pr. år. På grunn av drifts- og innkjøringsproblemer var rapporterte utslipp i 2003 og 2004 igjen tilbake på samme nivå som i perioden 1993-2000, med hhv. 0.37 og 0.31 g, men det skal nå igjen være nede på samme lave nivået som i 2001-2002, etter at det er gjort korrektive tiltak (pers. medd. Harald Eik, Avdelingsleder HMS & Kvalitet ved Falconbridge Nikkelverk AS).

**Tabell 1.** Oversikt over rapporterte utslipp av dioksin fra Falconbridge Nikkelverk 1988-2004

År	Utslipp g/år	Kilde for utslippstall
1988	5 – 6 [4.8]	Knutzen et al. (1998b) [tall i hakeparentes er lest ut av figur 1 nedenfor]
1989	5 – 6 [2.6]	
1990		
1991	0.6 [0.2]	
1992	1.6 [3.0]	SFTs oversikt over bedriftsspesifikk miljøinformasjon <a href="http://www.sft.no/bmi/">http://www.sft.no/bmi/</a>
1993	0.3	
1994	0.23	
1995	0.20	
1996	0.25	
1997	0.36	
1998	0.49	
1999	0.36	
2000	0.26	
2001	0.04	
2002	0.03	← Installert nytt renseanlegg
2003	0.37	
2004	0.31	



**Figur 1.** Tidsutvikling av dioksinutslipp til sjø hentet fra rapporten ”Miljøfakta 2004” fra Falconbridge Nikkelverk AS (<http://www.falconbridge.no>).

Analysar av biota i tidsrommet 1982-1996 har vist dioksin-nivåer som tilsier kostholdsråd og omsetningsrestriksjoner på sjømat fanget i området, selv om det ble påvist en nedgang i dioksin-konsentrasjonene i fisk ved siste prøvetaking i 1996 (Knutzen et al. 1998b). **Tabell 2** viser gjennomsnitsverdi og maksimum for TE for dioksiner i ulike prøvetyper fra Kristiansands-fjorden ved siste prøvetaking sammenlignet med det som er brukt som 'høy bakgrunnskonsentrasjon' ved overvåkingen av Grenlandsområdet (Knutzen et al. 2001). Høy bakgrunnskonsentrasjon er brukt som øvre grenseverdi for klasse I etter SFTs klassifiseringssystem, som beskrevet i Knutzen et al. (2001). Tabellen viser at det i 1996 forekom verdier opp til 15 ganger høy bakgrunnskonsentrasjon. Hvis dioksinkonsentrasjonene i Kristiansandsfjorden etter dette har sunket med 5-10 % pr. år slik en har sett i Grenland (Bjerkeng og Ruus 2002 og kapittel 5 til 8 nedenfor) etter at utslippene ble kraftig redusert, skulle konsentrasjonene i 2006 være nede i mellom 60 og 35 % av hva de var i 1996. Det indikerer at konsentrasjonene for torskelerver nå kan være kommet ned mot referansekonsentrasjonen, mens det for de andre prøvetypene antagelig fortsatt vil forekomme høyere konsentrasjoner. Siden nivåene er usikre pga. lite data, og det er usikkert om nedgangen har fortsatt med samme relative reduksjonsrate, er dette bare usikre antydninger om hva nivåene nå kan være.

**Tabell 2.** Referansekonsentrasjoner for dioksin i ulike vevstyper (Knutzen et al. 2001) sammenlignet med observerte konsentrasjoner i Kristiansandsfjorden ved siste overvåking. Konsentrasjoner er oppgitt som toksitetskvivalenter (TE) av dioksin i ng/kg våtvekt.

Prøvetype	Referanse konsentrasjon	Målt i Kristiansands-fjorden		Årstall
		Gjennomsnitt	Maksimum	
Blåskjell	0.2	1.5	2.4	1996
		0.52 (enkelt prøve etter to år med lave utslipp)		2002
Torskelerver	10	13	30	1996
Skrubbe filet	0.1	0.8	1.4	1996
Krabbesmør	10	36	96	1996

Det gjennomføres nå tiltak for å redusere sekundær belastning fra forurensede sedimentter, bl.a. ved overdekning. Det er da viktig med videre overvåking for å se hvilken virkning tiltakene har hatt, revurdere kostholdsråd, og gi prognoser for fremtidig utvikling. Dioksinanalyser er kostbare, og det er derfor spesielt viktig å optimalisere overvåkingen og vurdere hvilket nivå som kreves for å påvise endringer med ønsket grad av sikkerhet. Formålet med prosjektet er å gi en realistisk vurdering av hva en kan oppnå ved ulike prøvetakingsstrategier og se om det er mulig å gi råd om mest mulig kostnads-effektiv overvåking av fremtidig utvikling i dioksin-nivåer i Kristiansands-fjorden. Dette vil gi grunnlag for planlegging av videre overvåking.

I denne rapporten brukes eksisterende data for dioksin i biota fra Kristiansands-fjorden og Grenland til å estimere statistisk usikkerhet ved prøvetaking for å bestemme gjennomsnitskonsentrasjoner og ending over tid. På basis av dette vurderes hvordan et overvåkningsprogram bør gjennomføres for å oppnå rimelig usagskraft, dvs. mulighet til å påvise endringer.

## 2 Problemstilling

Hvis det skjer en langsiktig endring i eksponering for miljøgifter over tid, vil nivåene i organismer i gjennomsnitt reflektere dette, men de målte nivåene kan også variere ganske mye av andre årsaker, mer eller mindre uregelmessig ("tilfeldig"). Slik variasjon kan skyldes biologiske forskjeller mellom enkelt-organismer eller del-populasjoner, eller ha sammenheng med uregelmessig variasjon i ytre miljøforhold (f.eks. vannsirkulasjon, oppvirvling og omrøring av sediment, fødetilgang) som en ikke kan ha oversikt over.

Variasjon i ytre miljøforhold over tid kan forårsake endret miljøgiftkonsentrasjoner i organismene, både ved å endre eksponeringen for miljøgifter, f.eks. konsentrasjon av miljøgifter i føde, eller ved at det endrer den biologiske tilstanden hos organismene som studeres. Miljøgiftkonsentrasjonene kan også tenkes å påvirkes av svingninger i populasjonsdynamikk hos de populasjonene som studeres. slike variasjoner kan være uregelmessige på kort sikt, som mer eller mindre regelmessige sesongvariasjoner, men det kan også være variasjoner fra år til år eller på lengre sikt.

I tillegg til disse faktorene, som kan gi variasjoner i miljøgiftinnhold i organismer på litt større skala i rom og tid, vil det være en variasjon på liten skala, f.eks. mellom individer, som arter seg som en tilfeldig utvalgsusikkerhet. Det kan også skje endringer over tid i prøvetakingsmetoder eller tilgang på prøvemateriale. Det kan f.eks. tenkes at det ikke alltid er mulig å finne tilsvarende prøvemateriale på samme sted hvert år.

Selv om mye av påvirkningen i perioden etter store reduksjoner av utslippene kan antas å skyldes sekundære kilder, dvs. rester av tidligere utslipp, kan det selsagt også være variasjoner fra år til år i fortsatte direkte utslipp som påvirker nivåene i organismer merkbart.

De målte miljøgiftkonsentrasjonene i organismer kan altså variere over tid rundt en langsiktig jevn tidsutvikling (trend), evt. rundt et konstant nivå, av mange årsaker som kan gripe inn i hverandre og samvirke. Mye av disse variasjonene har en uforutsigelig, tilfeldig karakter, og utgjør derfor en "støy" i dataene. Den kvantifiseres ved *variansen*, som er middelveiden av kvadratet av avvikene fra trender eller gjennomsnitt. Standardavviket, som er kvadratroten av variansen, er det vanlige målet for et 'typisk' avvik,. Støyen gjør det vanskeligere å oppdage og kvantifisere langsiktige endringer over tid, fordi det gir tilfeldige utslag i middelveider over lengre tid og i koefisientene for tidsvariasjon som beregnes av observasjoner. De 'tilfeldige' utslagene i beregnede trender blir i alminnelighet mindre jo mer data en har å bygge på for å beregne en tidstrend, og jo lengre tidsperiode en kan se på, men vil alltid være der som en usikkerhet. Som nevnt kan svingninger også gjøre seg gjeldende over flere år, og med prøvetaking i en begrenset periode kan det da være vanskelig å skille fra hverandre endringer som skyldes eksponering og hvilke som har andre årsaker. Ved vurdering av forskjeller mellom geografiske områder kan en justere for variasjoner i tid som er felles for områdene, mens uavhengige variasjoner i hvert område vil gi "støy" i beregning av gjennomsnittlige forskjeller.

Ved statistisk testing av mulige effekter, f.eks. endringer over tid eller forskjeller mellom områder, sammenlignes beregnet effekt med det som kan forventes som en restvirkning av de tilfeldige avvikene. Hvis effekten er innenfor det som kan forventes tilfeldig, blir konklusjonen at en ikke kan påvise noen effekt. Hvis effekten er så stor at det er lite sannsynlig at den bare skyldes en resteffekt av de tilfeldige utslagene, konkluderes det med at det er en effekt. En må da veie mot hverandre risikoen for å ta feil i den ene eller andre retningen (se Vedlegg C for en mer detaljert beskrivelse).

Ved å estimere variansen for tilfeldige utslag ut fra eksisterende data kan det vurderes på forhånd hvor mye data en trenger fra ny overvåking for å påvise effekter av en viss størtelse med rimelig sikkerhet, og hvordan prøvetakingen bør fordeles i tid og rom og på ulike prøvetyper. Det er et mål å finne måter å estimere langsiktige endringer eller permanente forskjeller med minst mulig forstyrrelse av de tilfeldige utslagene. Hvis prøvetakingen kan gjøres på en slik måte at de tilfeldige variasjonene er små

eller utføves effektivt, vil effekten komme fram tydeligere, og testen blir mer følsom for små effekter. Det oppnås først og fremst ved å utvide prøvematerialet, dvs. ta flere prøver eller ta prøver over lengre tid. Noe kan også tenkes å oppnås ved å velge et optimalt sett av målevariable, i dette tilfelle dioksin-komponenter. En kan også prøve å korrigere måleverdiene for ulike faktorer hvor det er en kjent sammenheng (fysiologi, årstid), velge prøvemateriale (arter, vev) som har lite tilfeldig variasjon mellom enkeltprøver fra samme tid og sted i forhold til responsen på ytre eksponering av miljøgifter, og ved å fordele prøvetakingen over tid og sted innenfor et år slik at de uregelmessige variasjonene fra år til år blir jevnet ut best mulig. På den måten kan utsgnskraften for en prøveserie over et gitt antall år bedres.

Det er uregelmessig variasjon både mellom individer fra samme tid og sted og i gjennomsnitsnivåer for ulike prøvetakinger, dvs. forskjellige tidspunkt og/eller steder. Den siste typen variasjon kan være uavhengige utslag fra sted til sted ved samme tidspunkt, eller parallelle utslag over flere stasjoner i et område. Ved å anslå hvor store disse forskjellige bidragene til variasjon er, og analysere hvordan de bidrar til variasjon i middelerterter fra år til år, kan en si noe om hvordan utsgnskraften varierer med overvåkningsprogrammets omfang, hyppighet og varighet. Dette kan gjøre det mulig å trekke konklusjoner om hva som er et optimalt overvåkningsprogram ut fra hvor små endringer en er interessert i å påvise i gitt tidsperspektiv, sett i sammenheng med kostnader.

I de etterfølgende kapitlene er det gjort slike vurderinger for overvåkning av dioksin i organismer i Kristiansandsfjorden basert på materiale både derfra og fra Grenland.

## 3 Beskrivelse av datamateriale

### 3.1 Utvalg av datamateriale – geografiske områder og vevstyper

Den statistiske analysen baserer seg på data fra overvåkningen av miljøgiftinnhold i organismer i Kristiansandsfjorden og Grenlandsfjordene. For disse områdene er valgt ut data for blåskjell, torskelever, skrubbefilet og krabbesmør. Det er for disse prøvetypene det finnes data i et slikt omfang at det er mulig å estimere varianskomponenter til bruk i en analyse av utsgnskraft av måleprogrammer. Så og si alle data gjelder blandprøver av flere individer; antall individer er registrert for de fleste prøvene. For torsk finnes også noe data for filet, analysert på blandprøver fra samme individer som i noen leverprøver, men det er ikke tatt med her. For krabbe finnes også data for klokjøtt og for innmat, men bare data for krabbesmør er tatt med her. Dette utvalgte materialet omfatter 63 av i alt 89 prøveresultater fra Kristiansands-fjorden, og 246 av i alt 416 prøveresultater fra Grenland. De kjemiske analysene er gjort av Institutt for Folkehelse (IFFH) og Norsk institutt for luftforskning (NILU). Noen få av resultatene er parallelle analyser gjort ved IFFH og NILU på samme prøvemateriale, men de fleste prøvene er bare analysert av ett laboratorium. Det utvalgte materialet omfatter bare data for prøver som er funnet beskrevet i NIVA-rapporter eller stammer fra det nasjonale overvåkningsprogrammet (JAMP).

Alle data er registrert på våtvektsbasis. For omregning til annen basis er tørrvektspersent bare oppgitt for en del av prøvene, mens fettprosent er registrert for de fleste prøvene (se mer om dette i kapittel 3.4).

### 3.2 Geografisk og tidsmessig dekning i valgte områder

Prøvene er stedfestet både ved delområde og stasjonsnavn. For Grenlandsfjordene gir delområde og stasjon i stor grad samme oppdeling, i alle fall hvis en begrenser utvalget til stasjoner med data fra flere år. Stort sett er de statistiske analysene av Grenlandsdata basert på oppdeling i delområder. Også for Kristiansandsfjorden er det valgt å basere geografisk oppdeling på delområde, og med område-betegnelser A-H fra overvåkningsrapportene. Områdene B/C er slått sammen til ett område ut fra

tidligere vurderinger (Knutzen m.fl. 1991, 1998). Datamaterialet for Kristiansandsfjorden er for sparsomt til å etterprøve om det er signifikante forskjeller mellom stasjoner innenfor et delområde. Noen stasjoner eller områder har så lite data at de ikke kan inkluderes i en statistisk analyse av variasjon over tid og sted. Dette er redegjort for i sammenheng med de ulike delanalysene som presenteres nedenfor.

De fleste prøvene er tatt enten om våren i perioden mars-mai, eller om høsten i september – oktober, men det er også noen prøver fra andre måneder. Blåskjell, torsk og skrubbe er stort sett tatt om høsten i Kristiansandsfjorden, mens de fleste prøvene fra Grenland er tatt om våren. Krabbe er bare fanget om høsten i begge områder. For analyse av variasjon i tid er det i noen sammenhenger brukt overvåkingsår som går fra juli til juni neste år, slik det er brukt i JAMP-programmet. Der det ikke sies noe om dette gjelder tidsangivelsen kalenderår.

En mer detaljert oversikt for de to områdene er gitt nedenfor.

### 3.2.1 Grenland

Fra overvåkningen av Grenland finnes data for mer enn 400 dioksinanalyser, av disse er 246 valgt ut for nærmere analyse. Data spenner over tidsrommet 1976-2003. For blåskjell finnes ganske komplette tidsserier for 1987-2003 fra tre stasjoner. For torsk er det lengre tidsserier fra tre områder, med ganske komplett dekning for 1991-2002. For skrubbeflyndre finnes to kortere tidsserier med noen hull og begrenset grad av overlapping i tid mellom seriene. For krabbe finnes det flere serier som dekker tidsrommet 1987-2002 med litt varierende dekning av starten og slutten av dette tidsrommet.

**Tabell 3.** Oversikt over geografisk og tidsmessig dekning av utvalgte dioksindata i biota fra Grenland

Område	Blåskjell	Torsk	Skrubbeflyndre	Krabbe, ca. 70 % av årene i tidsrommet, varierende.
Frierfjorden	Steinholmen: Prøver i 1996 og 2002	1975-2003, alle år siden 1991	1987-1993, 1996 og 2000	1987-2000
Breviksfjorden	Croftholmen: Langtidsserie 1987-2003	1988-2001 og 2003	Alle år 1991-1997, samt 2000	1990-2002
Eidangerfjorden		2000 og 2002		
Langesundsområdet	Helgeroa: Langtidsserie 1987-2003	alle år 1988-2003		Fra tre steder: 1987-2001
Jomfrulandområdet:	Klokkartangen: 1987-2003, ujevn dekning, ca. halvparten av årene.			1988-2002

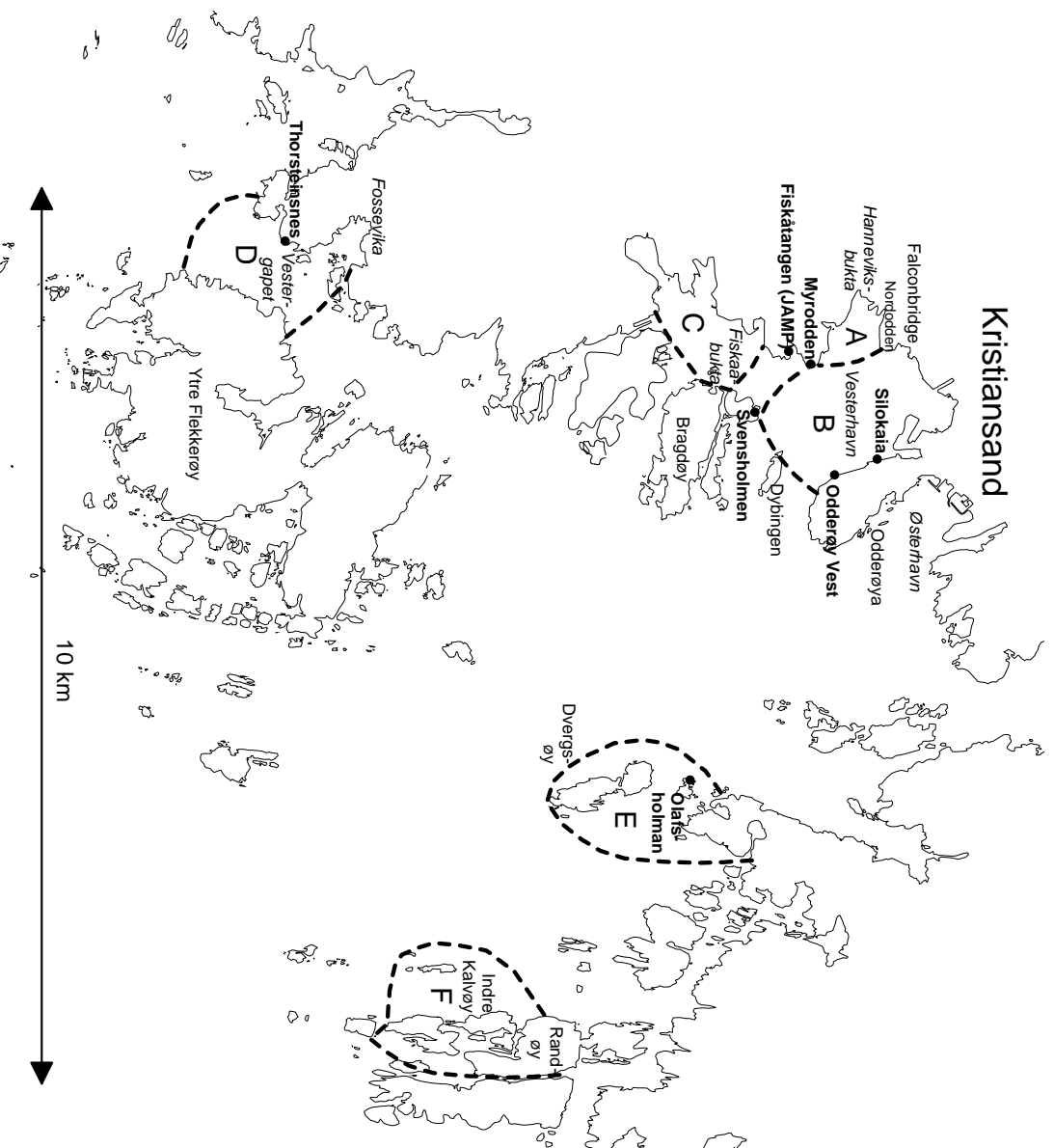
### 3.2.2 Kristiansandsfjorden

I alt er det samlet dioksindata for 86 biota-prøver fra Kristiansandsfjorden, av disse utnyttes 63 utvalgte prøver i denne analysen. Mest data er det for torsk, skrubbeflyndre, krabbe og blåskjell, og det er for disse fire artene det er aktuelt å gjøre artsspesifikke vurderinger. I tillegg er det data fra noen få prøver av sjøørret, rødspette, ål og reker, men det er for lite til å kunne analysere varianskomponenter.

Data for disse artene vil derfor bare kunne brukes i begrenset grad, for problemstillinger som kan undersøkes uavhengig av art.

Data dekker stort sett tidsrommet 1982-2002, og det aller meste er fra årene 1988-1996. Det har i hovedsak vært prøvetaking omtrent annet hvert år i den sistnevnte perioden. Det har vært varierende dekning av arter og geografiske områder fra år til år.

For fisk og krabbe er gjort fangster innenfor avgrensede områder. For blåskjell er det tatt prøver i noen definerte punkter. Delområdene og stasjonene er vist i **Figur 2** og prøvetaking er beskrevet i **Tabell 4**.



**Figur 2.** Kart over Kristiansandsfjorden med stasjoner for blåskjell (●) og fangstområder for krabbe og fisk (stiplet omriss), etter Knutzen et al. 1986a, 1986b, 1988a, 1991b, 1994a, 1998b og Norman Green pers. medd.

**Tabell 4.** Definerne delområder for prøvetaking av blåskjell, fisk og krabbe i Kristiansandsfjorden, med angivelse av når det er tatt prøver, angitt som enkeltår eller tidsrom med prøvetaking annethvert år. Blåskjellprøver fra 2002 er tatt innenfor JAMP-programmet og ligger alle i tilknytning til delområde B, se nærmere beskrivelse i kap. 5.2.1.

Del-område	Beskrivelse av området:	Blåskjell	Torsk	Skrubbe	Krabbe
A	Hanneviksbukta, vest for linjen Myrodden – Nordodden (østenden av Falconbridgekaia).	1986, 1990, 1995, 1996	--	1984, 1986	--
B	Vesterhavna øst for område A, avgrenset av en linje fra Myrodden via Dybingen til sørspissen av Odderøya.	1988, 1990, 1992, 1995, 1996, 2002	1984-1992, 1996	1982, 1988-1992, 1996	1988-1992, 1996
C	Fiskåbukta.	--		1988	--
D	Vestergapet, vest for Ytre Flekkerøy.	1988	1988, 1992, 1996	1996	1998, 1992, 1996
E	Området rundt Dvergøy på østsiden av fjorden.	1988	1988, 1992, 1996	--	1998, 1996
F	Området rundt Indre Kalvøy lenger sørøst i fjorden.	--	1988-1992, 1996	1988-1990	1988, 1990
G	Ny Hellesund (referansestasjon), ca. 9 km vest for Kristiansands-fjorden.	1988	1988, 1996		1988, 1996

3.3 Data om andre miljøgifter som støttevariable

For mange av prøvene er andre klororganiske stoffer, analysert ved NIVA, også tatt med i data-materialet, siden det var aktuelt å se på om overvåking av dioksiner kan gjøres mer kostnadseffektiv ved å utnytte sammenheng med konsentrasjon av andre stoffer som er billigere å analysere. Som det fremgår av analysen senere ble det ikke funnet grunnlag for å trekke dette inn i vurdering av utsagnskraft som noe som kan gi grunnlag for å minske overvåkningskostnader, men det er likevel nevnt, siden det kan tenkes å bli utnyttet ved en senere statistisk analyse til å øke utsagnskraften, dersom data foreligger (selv om vi ikke har kunnet påvise sammenhenger her, kan det tenkes at de vil fremkomme ved en grundigere analyse med mer data).

Ideen med dette er at hvis det er god korrelasjon mellom avvik fra gjennomsnitt eller trender for dioksiner og andre klororganiske (HCB, OCS, PCBer), kan overvåkningsprogrammet designes slik at et bare noen få prøver analyseres på både dioksiner og andre stoffer. Et større antall prøver kan analyseres bare på de andre stoffene, som har lavere analysekostnad. Målte dioksin-nivåer kan da justeres ut fra forholdet mellom nivåer i de andre stoffene i de to gruppene av prøver.

Anta som et eksempel at det ett år er analysert på miljøgifter i flere prøver fra et område, men på dioksiner bare i én prøve. Hvis denne prøven da har signifikant høyere nivåer for de andre miljøgifter enn gjennomsnittet over de andre prøvene, og det er påvist en korrelasjon mellom dioksin og de andre miljøgiftene, kan det antas at også dioksinkonsentrasjonene i denne prøven er høyere enn hva gjennomsnittet ville vært hvis hele materialet var analysert på dioksiner. Ved å justere ned dioksin-

konsentrasjonene ut fra dette kan en da kanskje få en tidsserie med mindre "støy". Dette forutsetter at småskala-variasjonene for dioksiner og de andre stoffene viser en generell korrelasjon (dvs. at variasjon mellom individer eller mellom nærliggende prøve-steder eller -tider viser samme mønster for dioksiner og andre stoffer).

Miljøgiftvariable som det er aktuelt å se på som støttevariable i overvåkning av dioksin-nivåer er først og fremst HCB, OCS og DCB (=CB209). De finnes sammenstilt med dioksinresultatene for ca. 2/3 av prøvene i begge områder<sup>1</sup>, og har også relativt god detekterbarhet. Det gjelder spesielt HCB. For DCB er ca. 10-15 % av prøvene under kvantifiseringsgrensen, og for OCS gjelder det 20-30 % av prøvene, slik at disse komponentene ikke er så verdifulle i denne sammenhengen. PCB-komponenter CB101, CB118, CB138 og CB153 ser også ut til å ha relativt god detekterbarhet, men finnes sammenstilt med dioksindata for langt færre prøver, slik at de ikke kan utnyttes så godt i denne sammenhengen.

For Kristiansandsfjorden finnes QCB(=5CB), HCB, OCS, DCB og Sum PCB for de fleste prøvene fra 1988 og 1990, og for alle prøver fra 1992 OG 1996. Prøvene fra 1996 har dessuten data for PCB-komponenter, CB118, CB138 og CB153 er de det finnes mest av.

### 3.4 Biologisk karakteristikk av prøvene

I den grad sammenhenger mellom dioksinkonsentrasjoner og biologisk karakteristikk kan påvises, kan det tenkes brukt til å korrigere eller normalisere de målte nivåene og derved minske støyen i data ved bestemmelse av tidstrender og geografiske forskjeller. For fisk er gjennomsnittlig lengde og vekt for de individene som inngår gitt for de fleste prøvene..

Det er også aktuelt å se på hvilken basis (våttvekt, tørrvekt, fettvekt) som er mest egnet. Alle data er som nevnt registrert på våtvektsbasis. Tørrvektspresent er bare oppgitt for en del av prøvene og kan vanskelig utnyttes. Fettprosent er bestemt for de fleste prøvene, stort sett av NILU eller IFFH samtidig med dioksinanalysen, evt. også av NIVA ved analyse av andre klororganiske stoffer. I den statistiske analysen er brukt fett-prosent bestemt av laboratoriet som har gjort dioksinanalysen hvis det er tilgjengelig; hvis ikke brukes fettbestemmelsen fra enten NILU, NIVA eller IFFH i nevnte prioritetsrekkefølge. Fettprosent mangler helt for Grenlandsdata fra 1976, 1987 og delvis 1988 og for en enkelt blåskjellprøve fra 1990, for data fra Kristiansands-fjorden fra 1982 og 1984 og dessuten for noen få av prøvene fra JAMP-programmet.

---

<sup>1</sup> Det finnes for flere prøver fra Grenland, men det gjenstår noe arbeid med å legge inn alt som finnes i databasen.



### 3.5 Avvikende verdier – utlukning av mulige feil.

Generelt er det ganske høy grad av samvariasjon mellom de forskjellige dioksinkomponentene. Ved å inspisere data i matriseplott<sup>1</sup> er det funnet en del avvikere fra dette. I en statistisk analyse for å påvise trender og forskjeller kan det være hensiktsmessig å se bort fra de mest avvikende verdiene for å frem hovedtrenderne bedre, dersom estimerte usikkerheter justeres for dette. Målet med analysen her er imidlertid å estimere hvor stort prøvetakingsprogram som kreves for å påvise endringer i en slik senere analyse, og da er det ikke riktig å se bort fra avvikende data generelt – vi kan ikke regne med tilsvarende mønster av avvikere for senere innsamlet materiale. Det er likevel noen få avvikere som er så markerte at det er valgt å ekskludere dem som sannsynlige feil:

- Torskelever, analyse av enkelindiivid, Frierfjorden 1987 (NILU prøvenr. 8703/2586): OCDF og 1234678-HpCDD er angitt som under deteksjonsgrense på hhv. 0.05 og 0.04 pg/g våtvekt, det er minst 1000 ganger mindre enn forventet ut fra resten av dioksiprofilen.
- Blåskjell, Breviksfjorden, Croftholmen, 1992: (NILU Prøvenr. 92/32): 1234678 HpCDD er ca. 1000 ganger lavere enn forventet ut fra resten av dioksiprofilen.
- Krabbesmør, Grenland, Ringsholmen 2000 (NILU Prøvenr. 00/1379): OCDF  $< 0.5$ ; det er i alle fall 100 ganger lavere enn forventet ut fra resten av dioksiprofilen. Verdien er dessuten markert i analyserapporten fra NILU med (g) som angir at gjenvinningsgraden ikke tilfredsstiller analysestandard.
- 

### 3.6 Detekterbarhet for dioksinkomponenter – vurdering av hvilke enkelt-komponenter og kombinasjoner som egner seg best for statistisk analyse.

Datamaterialet omfatter for omtrent alle prøvene et bestemt sett av komponenter, de såkalte 2378-komponenter (etter posisjonen til kloratomer i molekylene). I tillegg er det for ca. halvparten av prøvene også oppgitt sum-konsentrasjoner etter grad av klorering, f.eks. Sum-HpCDF (Sum Heptakloridbenzofuran). Summene representerer en automatisk generert tilleggsinformasjon som ikke er kvalitetssikret på samme måte som enkeltkomponentene; de er ikke med i de eldste prøvene og er heller ikke inkludert i de nyeste prøveresultatene. Statistisk analyse anbefales basert på data for enkelt-komponentene (Martin Schlabbach, NILU, pers. medd.). I det følgende betraktes derfor bare enkelt-komponentene for seg, eller kombinasjoner beregnet direkte av enkeltkomponenter. Toksitetets-ekvivalenten som er en vektet sum av enkeltkomponenter er et eksempel på dette (definert i **Tabell 5**). De sumvariable som er gitt som tilleggsopplysninger fra analysene brukes derimot ikke.

For å finne ut hvilke komponenter eller kombinasjoner som har best detekterbarhet, og som derfor er mest egnet til statistisk analyse, er det laget en sammenstilling av gjennomsnittlig konsentrasjon og bidrag til TE fra ulike komponenter, med oppdeling av hvor mange prøver som har verdi under deteksjonsgrense eller av annen grunn ikke er kvantifiserbare, og hvor mange verdier som er markert som spesielt usikre i analyserapportene. Sammenstillingen er gjort for de utvalgte prøvene fra Kristiansandsfjorden og Grenland hver for seg. Resultatet for de komponentene som er inkludert i beregning av toksitetsekvivalenten vises i **Tabell 5** nedenfor.

---

<sup>1</sup> XY-Plott av variablene parvis mot hverandre, satt opp i en matrise slik at en kan se dem i sammenheng. Det var logtransformerte verdier som ble plottet.

**Tabell 5.** Oversikt over gjennomsnittlig konsentrasjon og grad av detekterbarhet for dioksinkomponenter som inngår i beregning av toksisitetsekvivalent for utvalgt materiale fra Kristiansandsfjorden og Grenland.

Komponent	TE- Faktor	Uvalgte data fra Kristiansandsfjorden (63 prøver)					Uvalgte data fra Grenland (246 prøver)				
		Direkte målt	Bidrag til TE	kvan- tifi- sert	usikker verdi	Direkte målt	Bidrag til TE	kvan- tifi- sert	usikker verdi		
		Gjennomsnittlig konsentrasjon (ng/kg våtvekt)	observasjoner uten sikker verdi	Gjennomsnittlig konsentrasjon (ng/kg våtvekt)	observasjoner uten sikker verdi						
2378-TCDD	1	5.24	5.245	17	6	35.5	35.51	14	10		
12378-PeCDD	1	3.46	3.462	9	10	54.6	54.60	4	15		
123478-HxCDD	0.1	2.09	0.209	29	4	28.7	2.87	34	25		
123678-HxCDD	0.1	4.71	0.471	6	6	105.7	10.57	0	15		
123789-HxCDD	0.1	4.05	0.405	17	7	45.3	4.53	6	16		
1234678-HpCDD	0.01	5.83	0.058	3	6	47.1	0.47	1	12		
OCDD	0.0001	9.25	0.001	0	4	32.4	0.00	0	15		
2378-TCDF	0.1	60.79	6.079	0	0	420.3	42.03	0	17		
12378/12348-PeCDF	0.05	33.91	1.696	0	1	829.4	41.47	0	3		
23478-PeCDF	0.5	17.73	8.867	0	1	343.6	171.79	0	4		
123478/123479-HxCDF	0.1	23.17	2.317	0	1	1099.8	109.98	2	2		
123678-HxCDF	0.1	19.53	1.953	0	2	1184.1	118.41	2	4		
123789-HxCDF	0.1	4.14	0.414	22	4	66.1	6.61	13	24		
234678-HxCDF	0.1	8.08	0.808	2	3	122.3	12.23	0	9		
1234678-HpCDF	0.01	20.86	0.209	1	2	354.2	3.54	0	5		
1234789-HpCDF	0.01	5.12	0.051	11	8	81.2	0.81	3	6		
OCDF	0.0001	26.68	0.003	4	5	209.2	0.02	3	14		

Det dominerende bidraget til toksisitetsekvivalenten i begge datasettene er 23478-PeCDF. De 8 største bidragene til TE er markert med fet skrift for begge områder; det er de samme komponentene i begge geografiske områder. Komponenter som har ikke-detekterbare konsentrasjoner i mange prøver er skravert grått i de kolonnene som viser antallet ikke kvantifiserte.

6 komponenter skiller seg ut med høy konsentrasjon, høyt bidrag til TE og samtidig god detekterbarhet, dvs. ingen eller svært få prøver med ikke-kvantifiserbare mengder, i begge geografiske områder. Disse komponentene er 2378-TCDF, 12378/12348-PeCDF, 23478-PeCDF, 123478-123479-HxCDF og 123678-HxCDF og 234678-HxCDF. De første 5 er kvantifisert i alle prøvene fra Kristiansandsfjorden, og de første tre av disse komponentene er også kvantifisert i alle prøvene fra Grenland. 2378-TCDF har noe større innslag av usikre verdier i Grenland-data enn de andre komponentene i denne gruppen.

Nesten like god detekterbarhet har de 4 komponentene 1234678-HpCDD, OCDD, 1234678-HpCDF og OCDF, men de betyr mye mindre for toksisitetsekvivalenten. De to siste har også relativt høye konsentrasjoner. Disse komponentene kan derfor være verdifulle for statistiske analyser av trender selv om de ikke er så viktige i seg selv.

To komponenter, 2378-TCDD og 12378-PeCDD, har høyt bidrag til TE, men mye dårligere detekterbarhet. De må tas i betraktning ved vurdering av kostholdsråd og miljøeffekter ellers, men kan være litt vanskelige å bruke i vanlige statistiske analysemetoder for påvisning av tidstrender og geografiske forskjeller. I en faktisk analyse av tidstrender kunne det eventuelt vurderes å estimere dem ut fra sammenheng med andre komponenter i de prøvene hvor de ikke er kvantifisert, for å kunne utnytte den informasjonen de inneholder i andre prøver, men det ikke funnet nødvendig her, hvor formålet er å vurdere utsagnskraft. Endelig er det en gruppe komponenter som verken gir vesentlige bidrag til toksisitetsekvivalenten eller har god detekterbarhet. Det gjelder alle HxCDD, i tillegg til 123789-

HxCDF og 1234789-HpCDD. De komponentene som har relativt dårlig detekterbarhet i materialet fra Kristiansandsfjorden er vist i tabellen i kursiv med grå skrift i navnekolonnen.

I tillegg til de som er vist i tabellen er det noen spesifiserte komponenter som ikke finnes i noen data fra Kristiansandsfjorden, og for svært få prøver for andre områder, og da bare fram til 1990. De er heller ikke med i beregningen av toksisitetsekvivalenter. Det gjelder 12368/13479-CDF, 12478-CDF, 1278-CDF, 1378/1379-CDF og 2368-CDF. De er ikke vurdert her.

For å analysere usikkerheter og utsagnskraft og vurdere hvordan overvåkning bør legges opp er det ikke nødvendig å utnytte all informasjon; det er i stedet hensiktsmessig å konsentrere seg om de best detekterbare komponentene. Det ble derfor valgt å begrense analysen først og fremst til de 6 viktigst og mest målbare komponentene: 2378-TCDF, 12378/12348-PeCDF, 23478-PeCDF, 123478/123479-HxCDF, 123678-HxCDF, 234678-HxCDF, men også se på 4 andre komponenter: 1234678-HpCDD, OCDD, 1234678-HpCDF og OCDF der det var mulig. I tillegg til enkeltkomponentene er det sett på følgende typer kombinasjoner av flere dioksinkomponenter:

- Sum av TE-bidrag (TE<sub>n</sub>)  $\sum_{i=1}^n TE_{faktor_i} \cdot C_i$
- Direkte sum av konsentrasjoner (SUM<sub>n</sub>)  $\sum_{i=1}^n C_i$
- Geometrisk middelværdi (GM<sub>n</sub>)  $\sqrt[n]{\prod_{i=1}^n C_i} = \exp\left(\sum_{i=1}^n \log(C_i)\right)$

Det siste, geometrisk middelværdi, tilsvarer vanlig, aritmetisk middelværdi av log-transformerte verdier. Her bidrar komponentene til variasjon i middelværdi ut fra relativ variasjon, uansett absolutte konsentrasjoner.

Det er beregnet slike kombinasjoner for følgende sett av komponenter:

	TE3, Sum3, GM3	TE6, Sum6, GM6	TE10, Sum10, GM10
1234678-HpCDD			x
OCDD			x
2378-TCDF	x	x	x
12378/12348-PeCDF	x	x	x
23478-PeCDF	x	x	x
123478/123479-HxCDF		x	x
123678-HxCDF		x	x
234678-HxCDF		x	x
1234678-HpCDF			x
OCDF			x

Det ble beregnet to sett av verdier for hver kombinasjon: et nedre estimat hvor komponenter angitt som under påvisningsgrense (<verdi) eller hvor verdi mangler er satt =0 i kombinasjonen, og et øvre estimat hvor selve påvisningsgrensen er brukt der det er angitt at verdien er <påvisningsgrensen. Det øvre estimatet vil ha manglende verdi dersom noen av komponentene mangler. Underveis i analysen er alle disse kombinasjonene studert, men til slutt er det valgt å presentere resultater bare for de som bygger på de tre komponentene 2378-TCDF, 12378/12348-PeCDF og 23478-PeCDF når det gjelder å estimere varianskomponenter og vurdere teststyrke i et overvåkningsprogram. Ved senere statistiske analyser av overvåkningresultater kan det selvsagt være aktuelt å utnytte mer informasjon.

## 4 Metode

Kapittel 4.1 beskriver hvilke spørsmål om tidsutvikling en vil ønske å finne svar på med et nytt overvåkningsmateriale, dvs. hva slags effekter som det er aktuelt å teste. Det sies også noe om hvordan en kan vurdere utsagnskraft for disse testene.

Vurdering av utsagnskraft bygger på estimater for hvor stor uregelmessig variasjon det er i data-materialet, kvantifisert som midlere kvadratisk avvik, eller *varians*. Som diskutert i kapittel 2 er slik uregelmessig variasjon sammensatt av ulike bidrag med forskjellig tids- og rom-skala. Kapittel 4.2 beskriver i hovedtrekk hva slags analyser som vil brukes for å teste effekter, og hvordan eksisterende datamateriale analyseres med samme type analyser for å estimere varians for uregelmessig variasjon. Det er ment som en metodebeskrivelse av de analysene som er gjort i kapittel 5 til 8, hvor variansbidrag estimeres for forskjellige typer prøvemateriale.

Kapittel 4.3 diskuterer noen spørsmål som må tas i betraktning både ved vurdering av teststyrke og utsagnskraft og ved statistiske analyser for å teste hypoteser om utviklingen og er en bakgrunn for de vurderinger av utsagnskraft som er gjort i kapittel 9 på basis av variansestimaterne i kapittel 5 til 8.

### 4.1 Aktuelle statistiske tester på nytt overvåkningsmateriale

Tidligere prøvetaking i tidsrommet 1987 til 1996 viste konsentrasjoner i Kristiansandsfjorden som gjennomgående ligger over SFTS klassifiseringsgrense for klasse I, men at det var en nedgang over tid i tidsrommet. Dette blir nærmere analysert senere i rapporten. Ut fra dette er det naturlig å sette opp tre spørsmål en kan ønske å finne svar på for forskjellige områder og arter ved ny prøvetaking:

1. Har det vært ytterligere nedgang i konsentrasjonene siden 1996, og hvor stor har nedgangen vært?
2. Endres konsentrasjonene fortsatt over tid, og eventuelt hvor raskt?
3. Hva er konsentrasjonsnivået i forhold til SFTs klassifiseringsgrense?

For alle de tre spørsmålene vil det være naturlig å se på nye og gamle data i sammenheng, og bruke resultatene til å estimere nå-situasjonen og evt. predikere fremtidig utvikling ved fremskriving av observert tidsserie. Det kan bety at ending eller brudd i tidsutvikling bygges inn i modellen, men at alle data inngår i å beregne usikkerhet.

For de to siste spørsmålene kan det også være aktuelt å bare se på resultatene fra en ny overvåkning og vurdere resultatene uavhengig av eldre data. Siden det er gått 9 år siden forrige prøvetaking kan det være en brutt utvikling, f.eks. med utflåting av tidstrender, slik at det er mest naturlig å vurdere utviklingen nå bare ut fra nye data. Det vil spesielt være tilfelle dersom det er stor forskjell mellom gamle og nye data i størrelsen på fluktasjoner rundt en jevn tidstrend. Dersom det skulle vise seg at konsentrasjonene nå ikke endres vesentlig over tid, kan spørsmål 3 eventuelt vurderes ut fra enkel statistikk uten tidsutvikling, men mest sannsynlig vil det inngå som en del av vurdering av en tidstrend.

For å vurdere spørsmål 1 vil det være naturlig å sammenligne beste estimat for konsentrasjon i 1996 med middelverdien fra en ny overvåkningsperiode, og se om nivåene nå er signifikant lavere enn i 1996. Hvordan en skal estimere 1996-konsentrasjonene må vurderes ut fra en analyse av tidligere overvåkningsdata. Utsagnskraft vurderes ved å se på hvor stor nedgangen må være for at den skal påvises med rimelig høy sannsynlighet ved forskjellig omfang på ny overvåkning over 1 til 3 år, samtidig som vi har god sikkerhet mot å konkludere med at det har vært nedgang hvis det ikke er tilfelle.

For å vurdere spørsmål 2 må det gjøres en form for tidstrendanalyse. Målet for overvåking er å påvise om konsentrasjonene fortsatt øker eller avtar med tid. Det er mest aktuelt å se på utsagnskraft for å påvise reduksjon over tid. Nullhypotesen vil være at det ikke er noen reduksjon, og vi vil se hva som skal til for å forkaste denne hypotesen og konkludere med at det kan påvises en reduksjon. Vurdering av utsagnskraft baseres på at en tenker seg en jevn nedgang med en viss prosent av gjenværende konsentrasjon pr. år<sup>1</sup> og ser på hvor stor reduksjonstraten må være for at den skal påvises med rimelig sikkerhet ved forskjellige omfang og design av overvåkningsprogram. I dette inngår også å kreve god sikkerhet mot å konkludere med at det er en reduksjon dersom nivåene ikke reduseres over tid, men holder seg øker eller holder seg konstante.

For å vurdere spørsmål 3 vil vi sammenligne beregnet gjennomsnitt for dagens situasjon med en klassifiseringsgrense. Nullhypotesen vil være at konsentrasjonen ligger over grensen. Vurdering av utsagnskraft betyr da å se hvor langt under grensen den reelle middelkonsentrasjonen må ligge for at vi med god sikkerhet skal kunne konkludere med at den er innenfor grensen, samtidig som vi krever god sikkerhet mot å komme til feil konklusjon hvis konsentrasjonen faktisk ligger over grensen. Hvis den reelle konsentrasjonen ligger nær grensen i forhold til usikkerheten, blir konklusjonen mest sannsynlig at området ikke klassifiseres som innenfor grensen. For å modifisere dette kan en evt. legge inn en viss margin ved å justere opp den grensen som brukes i testen med en valgt, akseptabel toleranse.

## 4.2 Estimering av varianskomponenter

I kapitlene 5 til 8 er data for de forskjellige prøvetypene analysert for å beregne varianskomponenter til bruk i vurderingen av utsagnskraft i kapittel 9. Hver prøvetype er behandlet i et eget kapittel, hvor datamaterialet fra Grenland og Kristiansandsfjorden først er analysert hver for seg. Hvert kapittel avsluttes med en oppsummering hvor resultatene fra de to områdene ses samlet.

Det er valgt å konsentrere den statistiske analysen om de seks hovedkomponentene av dioksin som ble valgt ut i kapittel 3.6 og i tillegg de kombinasjonsmålene som er basert på tre dioksiner (TE3, SUM3 og GM3).

Alle analyser gjøres på logtransformerte<sup>2</sup> verdier (Vedlegg A). Ved å logtransformere konsentrasjoner kan forskjellige i ytre miljø og forskjellige i biologisk betingede faktorer skilles fra hverandre i additive statistiske modeller (med en sum av ulike ledd) selv om faktorene virker multiplikativt (f.eks. konsentrasjon i organismer = konsentrasjon i ytre miljø multiplisert med akkumuleringsfaktor).

Samtidig gir logtransformering ofte fordelinger som er mer symmetriske, og hvor variansen er mer uavhengig av gjennomsnittlig nivå slik de vanlige statistiske modellene krever. Det brukes naturlige logaritmer.

Analysene som gjøres er basert på såkalt *Generell Lineær Modellering* (GLM), som er en mer generell utgave av *varians- og kovariansanalyse*. Slike analyser anvendes for datasett hvor observerte verdier er antatt å variere med flere faktorer, som har effekt uavhengig av hverandre eller i kombinasjon.

Dette er samme type analyse som kan brukes senere for å teste hypoteser med et nytt overvåkningsmateriale. Ved hypotesetesting brukes varians/kovarians-analyse til å se om ulike effekter er

<sup>1</sup> Log-lineær regresjon av konsentrasjon ( $C$ ) med tid ( $t$ ):  $C = C_0 e^{-kt}$

<sup>2</sup> I denne rapporten er det konsekvent brukt naturlige logaritmer, som beskrevet i Vedlegg A. Det gjelder enten det er brukt betegnelsen  $\ln$  eller  $\log$ . Naturlig logaritme  $\ln(x)$  av et tall  $x$  er definert ved at  $x = e^{\ln(x)}$  hvor konstanten  $e = 2.7182$ . De naturlige logaritmene har den egenskapen at små forskjeller i logaritmene til to verdier gir direkte uttrykk for relativ forskjell, dvs. at  $\ln(x_1) - \ln(x_2) \equiv (x_1 - x_2)/x_2$ .

signifikante, dvs. større enn det som med rimelighet kan skyldes resteffekter av tilfeldige avvik. Da sammenlignes beregnet størrelse på effektene med sannsynlige utslag av variasjon som ikke beskrives som effekter i modellen (residualer) eller virkning av andre effekter som kommer inn som feilkilder.

Analysene i kapittel 5 til 8 gjøres altså på samme måte som når en vil teste effekter, men her er vekten lagt på å estimere varians for ulike bidrag til variasjon. De varianser som estimeres her vil ikke bli brukt ved senere testing av hypoteser på nytt eller utvidet datamateriale, men gir grunnlag for å vurdere hvor store tilfeldige utslag som kan opptre når en senere skal teste om effekter er signifikante. Ved slik testing blir varians beregnet på nytt ut fra det datamaterialet som inngår i testen.

Faktorene i en GLM-modell kan være *kategoriske*, hvor observasjonene klassifiseres i to eller flere grupper for hver faktor (stasjoner, arter, kjønn), eller *kovariate*, dvs. kontinuerlige variable (lengde, vekt, fett-%) som antas å være lineært korrelert med de måleverdier som studeres når andre ting holdes konstant. År kan behandles som kategorisk faktor eller som kovariat. Modellen kan inneholde en blanding av kategoriske faktorer og kovariater. Hvis modellen bare inneholder kategoriske faktorer, har vi en *variansanalyse*. Hvis den også omfatter kovariater, snakker vi om *kovariansanalyse*.

Data beskrives av modellen som en sum av ledd som beskriver virkningen av de ulike faktorene alene eller i interaksjon og ulike bidrag til tilfeldig variasjon. For det aktuelle tilfelle har vi data fra flere stasjoner over en serie år, og til dels med flere prøver (replikater) fra hver kombinasjon av stasjon og år. Ser vi bort fra eventuell korreksjon for biologiske karakteristikk, kan opdelingen av tilfeldig variasjon da skrives som vist nedenfor, hvor  $t$ = tid (som år),  $st$  = sted eller stasjon, og  $r$ =replikat eller prøvenummer innenfor gitt tid og sted:

$$\log(C_{t,st,r}^*) = f(t, st) + \underbrace{\varepsilon_t}_{\substack{\text{tilfeldig utslag} \\ \text{fra år til år} \\ \text{felles for} \\ \text{alle stasjoner}}} + \underbrace{\varepsilon_{t,st}}_{\substack{\text{tilfeldig variasjon} \\ \text{fra år til år} \\ \text{uavhengig} \\ \text{på hver stasjon}}} + \underbrace{\varepsilon_{r(t,st)}}_{\substack{\text{tilfeldig variasjon} \\ \text{mellom enkelprøver} \\ \text{innenfor samme} \\ \text{år og stasjon}}}$$

Funksjonen  $f(t, st)$  skal beskrive regelmessig og systematisk tidsutvikling og geografisk variasjon og har koeffisienter som beskrives ved tilpasning til data. De ulike  $\varepsilon$ -leddene utgjør 'residualer' i data sammenlignet med denne modellen og vil gi tilfeldige restvirkninger på effekter som testes. Målet med analysene i denne rapporten å estimere varians for de tre tilfeldige leddene ( $\sigma_t^2$ ,  $\sigma_{t,st}^2$ ,  $\sigma_r^2$ ) som grunnlag for å vurdere slike restvirkninger.

Variasjon mellom de enkelte år og interaksjon år\*sted er her satt opp som tilfeldige variasjoner og ikke bare som spesifikke effekter en vil teste om er til stede. Det er imidlertid også systematisk komponenter her, i alle fall i variasjon over tid, men kanskje også i interaksjon år\*sted ; de vil da inngå i funksjonen  $f(t, st)$ . Systematisk interaksjon år\*sted kan f.eks. være at konsentrasjonen reduseres raskere sett over lengre tid på noen stasjoner enn andre.

Analysene gjøres slik at en prøver å skille den systematiske variasjonen over tid og sted, dvs. jevne, langskitige endringer, fra de tilfeldige, uregelmessige variasjonene. Dette skillet innebærer at vi betrakter det enkelte år som et tilfeldig valgt prøvetidspunkt i et tidsforløp, eller det enkelte prøvested som tilfeldig valgt i et område, og at det er de generelle trekkene vi ønsker å si noe om. Minst en av faktorene *år* og *sted* må da inngå som tilfeldig faktor ('random factor') i analysen.

Når det gjelder tidsutviklingen kan det være aktuelt å teste på forskjeller mellom bestemte år, da betraktes år som fast faktor ('fixed factor') og  $\varepsilon_r$  vil inngå i effektene som testes, mens prøvested betraktes som tilfeldig faktor. Konstante forskjeller mellom stasjoner inngår i funksjonen  $f(t, st)$ , det tas med i analysene, men er ikke av primær interesse her.

For et datamateriale med flere prøver fra hver kombinasjon av år og sted kunne en også betrakte begge som faste faktorer, og teste signifikans av forskjeller mellom bestemte kombinasjoner av år og stasjon i forhold til utvalgssikkerhet for enkelprøve. Det kunne være aktuelt hvis en f.eks. ville se om variasjoner på en stasjon nær utslippet var systematisk forskjellig fra en stasjon lenger unna. De data som foreligger har så lite replikater at det uansett ikke er grunnlag for det.

Oppdelingen i regelmessig variasjon og ulike bidrag til tilfeldig variasjon er en forenklet beskrivelse av en kompleks virkelighet, og viser bare en av mange mulige måter å dele opp variasjonen på, valgt ut fra hva slags data som foreligger. Data er til dels fra varierende måned innenfor året gjennom en tidsserie; da kan variasjon mellom år og interaksjon år\*stasjon delvis være en virkning av systematisk variasjon mellom måneder, men dataene er for spredt til at det er mulig å skille det ut. Det er heller ikke grunnlag for å separere tilfeldig kortperiodisk variasjon innenfor året fra variasjon mellom år.

Modellen for tilfeldig variasjon viser hvilke ulike tilfeldige variasjonsbidrag som maksimalt kan skilles fra hverandre ut fra eksisterende data. Det er ikke mulig å få det til fullt ut for alle datasett som analyseres i etterfølgende kapitler. For de fleste datasett er det ingen eller svært få replikate prøver. Da er det ikke grunnlag for å skille de to siste leddene fra hverandre og interaksjonsleddet vi omfatte begge, og ha varians som er summen av variansen fra de to bidragene. Hvis data fra én stasjon analyseres for seg, kan en bare estimere samlet varians for summen av alle de tre leddene for tilfeldig variasjon. Har en ser på data fra to enkeltstående år, kan ikke den tilfeldige variasjonen mellom år skilles fra systematisk parallell variasjon over år i funksjonen  $f(t, st)$ , og de tilfeldige utslagene  $\varepsilon_{at}$  vil inngå i den effekten som estimeres.

Funksjonen  $f(t, st)$  kan ha ulike former, og hva som kan beskrives som residualvariasjon vil variere med formen på den systematiske variasjonen. Det er brukt litt forskjellige modeller for systematisk variasjon for å få frem variansestimater for tilfeldig variasjon. I alle analysene som gjøres her, er det lagt inn koeffisienter for forskjell i gjennomsnitt over tid mellom de enkelte delområder eller stasjoner. Når det gjelder tidsutviklingen er noen analyser gjort med en kovariansmodell som har lineære tidstrender på logskala, dvs. konstant prosentvis reduksjon pr. år av konsentrasjonen. År kan ikke samtidig inngå som en faktor med tilfeldig variasjon, og modellen blir:

$$\log(C_{t,st,r}^*) = \mu_0 + \alpha_{st} + b_{st} \cdot (t - t_0) + \varepsilon_{t,st} + \varepsilon_{r(t,st)}$$

I noen tilfeller er regresjonskoeffisienten  $b_{st}$  etter en vurdering av dataene antatt å være lik på alle stasjoner; eventuelle forskjeller mellom stasjoner i tidsutvikling vil da også fanges opp i leddet  $\varepsilon_{t,st}$ . Den tilfeldige variasjonen fra år til år vil som er felles for alle stasjoner ( $\varepsilon_t$ ) vil også inngå i den tilfeldige interaksjonen ( $\varepsilon_{t,st}$ ).<sup>1</sup> For å separere disse to tilfeldige bidragene er det i ett tilfelle gjort en sekundær variansanalyse på residualene fra kovariansmodellen.

Det er også brukt variansanalyser på dataene direkte, hvor tid (år) inngår som kategorisk faktor:

$$\log(C_{t,st,r}^*) = \mu + \alpha_{st} + \beta_t + \varepsilon_{t,st} + \varepsilon_{r(t,st)}$$

Da vil koeffisientene  $\beta_t$  fange opp både jevn tidsutvikling og flukturasjoner fra år til år, mens interaksjonen år\*stasjon ( $\varepsilon_{t,st}$ ) vil fange opp både uregelmessig variasjon og evt. forskjell i jevn

---

<sup>1</sup> Ved hypotesetesting må slik felles variasjon tas hensyn til, enklest ved at data fra alle stasjoner hvert år slås sammen til et gjennomsnitt for analysen av tidsserien av gjennomsnitt, se mer om det i Vedlegg B.

tidsutvikling. I et tilfelle er residualene fra en slik analyse etterpå analysert i en kovariansanalyse med lineær tidstrend, uavhengig på hver stasjon, for å estimere varians for den tilfeldig delen av  $\beta_i$  og  $\varepsilon_{i,st}$ .

Konstantene og koeffisientene som inngår for å beskrive forskjeller mellom grupper eller sammenheng med kovariater bestemmes tilnærmet (*estimeres*) ved tilpasning til et sett av observerte verdier. Det gjøres ved å velge verdier som gjør midlere kvadratisk gjennomsnitt (variansen) for de tilfeldige utslagene minst mulig. Den estimerte variansen for ulike bidrag til tilfeldig (uregelmessig) variasjon brukes så til å vurdere hvor store reelle effekter må være for at de skal bli påvist med rimelig sannsynlighet når samme type modeller senere brukes i statistiske tester. For en mer detaljert matematisk beskrivelse, se f.eks. Vedlegg B.

Det er i hovedsak tre typer analyser som gjøres i kapittel 5 til 8 for å estimere de ulike bidragene til tilfeldig variasjon:

- Enveis variansanalyse gjøres på data for replikate prøver fra samme tid og sted for de få prøvetakingstilfellene (sted\*tid) der dette forekommer. Kategorisk faktor er da kombinasjon tid\*sted. Hensikten er å estimere utvalgsusikkerheten pga. variasjon mellom individer eller mellom grupper på så liten skala at det gir tilfeldig variasjon mellom utvalgte deler av prøvematerialet. Ved å se hvordan denne variansen avhenger av antall individer i prøvene kan det også i noen tilfeller sies noe om hvor mye en kan vinne ved å ha flere individer i prøven.
- Toveis variansanalyse med år og sted (område eller stasjon) som faktorer gjøres på hele eller det meste av tilgjengelig datamateriale. I mange tilfelle er datamaterialet ufullstendig, dvs. at ikke det ikke er data fra alle kombinasjoner av år og sted. Ved noen av analysene ekskluderes enkeltstående prøver fra steder eller år med spesielt dårlig dekning for å bedre balansen i datamaterialet noe. Analysen vil generelt separere tre variansledd: mellom år som gjennomsnitt over steder, mellom steder i gjennomsnitt over år, og interaksjon sted\*år, dvs. uavhengig variasjon over tid på ulike steder. Variansen fra interaksjon sted\*år vil være uttrykk for fluktasjoner på hver stasjon som opptrer tilfeldig i forhold til tidspunkt for prøvetaking og med varierende varighet. At det opptrer som en effekt mellom år i dataanalysen skyldes frekvensen for prøvetakingen og ikke at det er et fenomen som opptrer på årsbasis. Det samme gjelder i prinsippet variasjonen mellom år som er felles over flere stasjoner. Interaksjonsleddet er også uttrykk variasjon på en geografisk skala som er definert av plasseringen av prøvetakingslokaliteter. Interaksjon bestemt for et sett av lokaliteter er derfor ikke nødvendigvis overførbar til andre områder. Til en viss grad kan den likevel være det, dersom det er en sammenheng mellom tidsskala og geografisk omfang for uregelmessige variasjoner.

- Som supplement til dette gjøres også kovariansanalyse med log-lineær regresjon mot år, og med sted som kategorisk faktor. I den toveis variansanalysen som er beskrevet ovenfor, vil varians mellom år i hovedsak være uttrykk for en tidstrend, og forskjell i tidstrend mellom steder vil inngå i interaksjonen sted\*år. I kovariansanalysen kan en legge inn separat trend for hvert sted (som interaksjon sted\*tid), og derved få bedre estimat for interaksjonsleddet.

Residualene fra kovariansanalysen analyseres etterpå i en ny toveis variansanalyse for å estimere variansledd for uregelmessig fluktasjon over tid som gjennomsnitt over steder og interaksjon tid\*sted. Resultatet vil være et supplement til resultatet fra variansanalysen. Det er valgt å bruke log-lineær regresjon som stemmer rimelig bra med observasjonene. Siden hovedhensikten er å estimere varians rundt tidstrenden, kan det være en fordel å bruke en relativt stiv modell for tidstrend som trekker i retning av noe for høye varianskomponenter; det vil ellers være en tendens til å undereestimere variansledd basert på lite data (Vedlegg F). I en statistisk analyse for å bestemme trender kunne det være aktuelt å bruke andre metoder, som ikke-lineær regresjon, glatting, eller ikke-parametriske metoder, men det anses ikke nødvendig her.



For hver prøvetype drøftes resultatene av de forskjellige analysene i sammenheng, både innen hvert geografiske område, og fra de to områdene sett samlet. Grenlandsmaterialet analyseres først, siden det er mye mer omfattende og gir derfor sikrere variansestimater enn data fra Kristiansandsfjorden.

Deretter gjøres en supplerende analyse av materialet fra Kristiansandsfjorden, som grunnlag for å vurdere om det er grunnlag for trekke generelle konklusjoner. Tilslutt gjøres en felles oppsummering av alle resultatene som utgangspunkt for vurderingen av overvåkningsprogram.

I forprosjektet ble det skissert en mulighet for å bruke analysemetoder som *PLS (Partial Least Squares)* og kanonisk korrelasjon for å lete etter kombinasjoner av dioksiner og aktuelle kovariater (område, lengde, vekt, levervekt, fett-%) som viser god sammenheng med hverandre og eventuelt kunne brukes til å gi mer stabile bilder av tidsutvikling i dioksineksponering korrigert for kovariater. Dette har vært prøvd i en viss grad, men det har ikke gitt meningsfylte sammenhenger – resultatene blir for tilfeldig styrt av enkelavvik i data. For eksempel kan en ende opp med en lineær kombinasjon av dioksiner hvor koeffisientene er en blanding av store positive og store negative koeffisienter, som ikke fanger opp generell forurensningsgrad, og ikke har noen fornuftig tolkning. I stedet er det altså lagt vekt på å se på dels de ulike enkeltkomponentene, og dels veldefinerte kombinasjoner som er definert på forhånd (summer, geometrisk middel). I forhold til den usikkerheten som ligger i variansestimaterne anses det ikke hensiktsmessig å lete mer etter avanserte måter å kombinere dioksinkomponentene på i denne rapporten.

Der det er grunnlag for det, er mulige sammenhenger med biologiske faktorer, som kjønn, lengde, og fett-% undersøkt. Det er funnet grunn til å justere for dette i noen analyser; i andre tilfeller er det funnet sammenhenger som er signifikante, men virker lite robuste, og da er det sett bort fra dem. Det vil være aktuelt å bruke dem i en analyse for å teste for forskjeller og tidstrender, men for å analysere teststyrke er det bedre å være konservativ og ikke basere analysen på usikre sammenhenger.

Det er også undersøkt om det er mulig å utnytte samvariasjon med andre miljøgifter som er rimeligere å analysere til å bedre teststyrken ved gitt kostnad. Inndertid er de sammenhengene som er funnet relativt svake, og de avhenger ofte av et fåtall avvikende prøver, slik at det er tvilsomt om sammenhengene er generelle. Av samme grunn som for biologisk karakteristikk er det derfor valgt å ikke basere analyse av teststyrke på slike sammenhenger.

### 4.3 Bruke resultatene til å vurdere utsagnskraft

Vedlegg B skisserer hvordan data fra prøvetaking på flere steder, gjentatt over et antall år, kan analyseres for tidstrender og nivåer, og hvordan utsagnskraften ved ulike tester vil avhenge av sammensetningen av overvåkningsprogrammet. Beskrivelsen er basert på lineær tidstrend (på logtransformerte data). Ved en senere analyse av data kan det tenkes at en vil bruke andre tidstrend-modeller, men for på vurdere utsagnskraft mot prøvetakingsprogram anses det tilstrekkelig å se på lineær-regresjon.

Utsagnskraften vil avhenge av hvilke fluktusjoner en har i data rundt de jevne trendene. Beskrivelsen har komponenter for fluktusjoner fra år til år, både som gjennomsnitt for hele området og uavhengig på hver stasjon. I tillegg er det lagt inn variasjon mellom replikate prøver tatt på samme tid og sted. Varians mellom replikate prøver kan være bare en utvalgsusikkerhet knyttet til variasjon mellom individer, eller også inneholde småskala variasjoner i tid og rom som det ikke er mulig å kontrollere eller ha oversikt over, f.eks. uregelmessige fluktusjoner fra uke til uke eller innenfor nøyaktigheten i posisjon for prøvetaking.

En slik beskrivelse er en forenklet strukturering av en variasjon pga. et mangfold av årsaker, over ulike avstands- og tidsskalaer, som går over i hverandre. Modellen splitter opp dette mangfoldet på en bestemt måte, tilpasset hvordan en planlegger og gjennomfører prøvetaking. De ulike leddene vil kunne inneholde ulike elementer avhengig av hvordan prøvetakingen gjøres, og de kan heller ikke alltid skilles fra hverandre.

---

Hensikten med overvåkingen vil være å estimere de generelle tidstrendene og gjennomsnitsnivåene, og eventuelt også variasjonen rundt den. For å vurdere krav til prøvetakingsprogram må vi bruke estimert varians for forskjellige typer tilfeldig variasjon, og se på hvordan virkningen av dem som usikkerhet ved test av effekter varierer med ulike overvåkningsprogram. For eksempel vil en ved å øke antall prøver pr. år kunne redusere restvariansen av både prøvetaking og eventuelle kortperiodiske variasjoner over tid innenfor et år.

For dioksindata er det i stor grad bare tatt én prøve pr stasjon og år. Da er det ikke mulig å skille fluktasjoner mellom år fra uregelmessig variasjon over korte tidsrom innenfor prøvetakingsperioden hvert år. Ved å tilpasse modellen til data vil vi estimere en residualvarians som er summen av variansen for de to leddene. For noen datatyper vil det finnes tilfeller av ekte replikater fra samme tid og sted, og da kan dette studeres separat og brukes sammen med andre variansestimater fra større datasett.

## 5 Statistisk analyse av data for blåskjell

### 5.1 Data fra Grenland

#### 5.1.1 Beskrivelse av data

Fra Grenlandsområdet er det sammenstilt data for dioksin i blåskjell for alle år fra 1987 til 2003, for det meste fra overvåkningsprogrammet for Grenland, men supplert med data fra JAMP-programmet. **Tabell 6** viser en oversikt over damaterialet. Prøvestedene er beskrevet i overvåkningsrapporter (Bakke et al. 2003, Green et al. 2001). Data for Gjermundsholmen fra JAMP-programmet er slått sammen med data fra den nærliggende Crofttholmen i Grenlandsovervåkingen. Det er særlig fra den stasjonen, og fra Helgeroa på østsiden av Langesundsfjorden lenger ut, at det er god dekning over tid. På Klokkartangen ned mot Jomfruland ble det tatt prøver de første årene, men etter 1991 bare enkelte år. For det meste er det 50 skjell i hver prøve, men i JAMP-materialet opp til 260 skjell i noen prøver.

**Tabell 6.** Oversikt over dioksindata for blåskjell fra Grenlandsområdet med fordeling på stasjoner og år. Tallet for hver kombinasjon av stasjon og år angir antall analyser.

Område	Stasjon	1987	'89	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	2000	'01	'02	'03
Breviks-fjorden	Frierfjorden									+1						+1	
	Crofttholmen	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Risøyodden		3 <sup>§</sup>							1+1 <sup>#</sup>						+4 <sup>§</sup>	
	Strømtangen															+1	
Langesundsområdet	Arøya						1										
	Helgeroa	1	1	1	1	1	1	1	1	1	2 <sup>#</sup>	1	1	1	1	1	1
Jomfruland	Klokkartangen	1	1	1	1	1	2 <sup>*</sup>				1			1		1	1

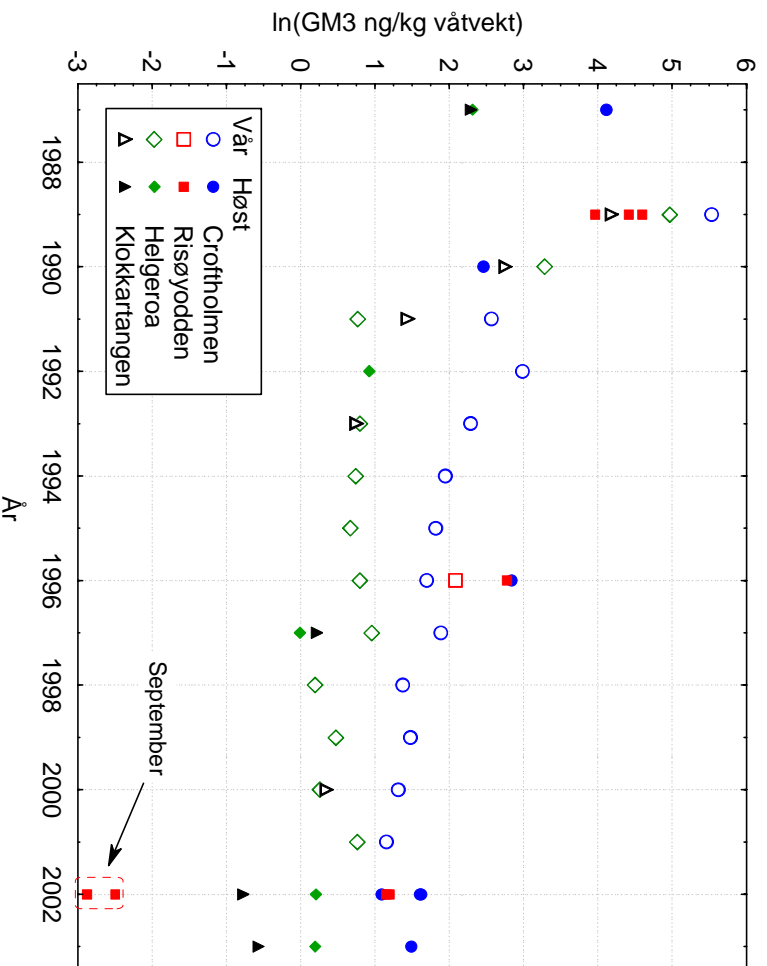
+n: Prøver fra JAMP-programmet  
\*: Parallell-analyse av samme prøve av NILU og IFTH.  
#: Analysert to prøver, en fra april, en fra august-september  
§: samtidige replikate prøver

Dette datamaterialet inneholder noen tilfeller av replikate prøver:

- På Risøyodden 1989 ble det tatt tre samtidige prøver på forskjellig substrat (på stein langt fra sediment, på fjell nær sediment, direkte på sediment).
- På Croffholmen i 2002 ble det tatt en prøve i september som ledd i overvåkningsprogrammet for Grenland, og 2 replikate prøver i oktober innen JAMP-programmet.
- På Risøyodden 2002 ble det tatt to par replikate prøver innen JAMP-programmet, ett par i september og ett i oktober.

De fleste prøvene er samlet fra april måned, men det er også noen fra mars eller mai, en fra juli, og noen fra september eller oktober (noen prøver 1996 og 1997, og fra 2002 og 2003). Det er tildels store forskjeller mellom replikater fra forskjellige måneder samme år, men ikke konsekvent i samme retning. Det er derfor tryggest å anta at variasjonsmønsteret i løpet av året ikke er forutsigbart, og ikke basere vurdering av utsagnskraft på å korrigere for måned.

For stasjonene Croffholmen og Helgeroa er det tatt prøver alle år innenfor perioden fra 1987 til 2003. også Klokkartangen har data fra flere år, men mer spredt i tid. Fra Risøyodden finnes data fra 1989, 1996 og 2002, hvert år med replikater. **Figur 3** viser eksisterende data for geometrisk middel av tre komponenter GM3, som definert i kapittel 3.6, på logskala. De andre komponentene og kombinasjonsmålene gir mer eller mindre det samme bildet. Åpne symboler er fra vårhåret, stort sett fra mars-mai, mens fylte symboler er fra høsten, vesentlig fra september til oktober.



**Figur 3.** Geometrisk middel av tre dioksinkomponenter (GM3, se kapittel 3.6) for utvalgte blåskjell-stasjoner i Grenlandsområdet for perioden 1991-2003.

De få tilfellene av prøver fra både vår og høst viser ingen tegn til systematisk årstidsvariasjon – forskjellige mellom vår og høst går begge veier. Sett i sammenheng med variasjon mellom replikater innenfor stasjon og år er det ikke grunnlag for å fastslå avvik fra log-lineært tidsforløp etter 1991 for de stasjonene som har mest data, men det er heller ikke noe klart bilde av en kontinuerlig tidsutvikling. Visuelt gir data inntrykket av en generell tendens til reduksjon i nivåene når en sammenligner data fra årene 1991-1997 med data fra 1998-2003, men innenfor den siste perioden er det ikke noen fortsatt generell nedgang for Croftholmen og Helgeroa, og for lite data fra Klokkartangen til å si klart om det har vært en nedgang. For Risøyodden er det svært lave konsentrasjoner i to prøver som ble innsamlet som del av JAMP-programmet i september 2002, mens prøven fra Risøyodden fra Grenlands-overvåkingen i oktober samme år fortsatt ligger på omtrent samme nivå som Croftholmen, slik det også gjorde i 1989 og i prøvene fra JAMP-programmet i 1996.

## 5.1.2 Statistiske analyser

Det gjøres forskjellige analyser på delvis overlappende datamateriale for å estimere varianskomponenter:

1. En variansanalyse er gjort på data fra Croftholmen, Helgeroa og Klokkartangen med år og stasjon som faktorer. Bare data fra og med 1991 er tatt med, fordi forholdet mellom stasjonene kan ha endret seg systematisk etter at utslippene ble redusert. År behandles som tilfeldig faktor, og interaksjonsleddet år\*stasjon er med som eget ledd i variansmodellen, skilt fra residualvariansen som estimeres ut fra de få tilfellene av mellom replikater innen stasjon og år. Denne analysen gir relativt høy residualvariens for replikaterne, ca. 0.33 for TE3, SUM3 og GM3, men bare bestemt med 4 frihetsgrader og derfor svært usikkert (80 % konfidensintervall 0.17 til 1.3). Residualvariansen omfatter her både variasjon mellom replikate prøver tatt på samme tid og mellom prøver tatt i forskjellige måneder innenfor samme år og fra samme stasjon. Midlere kvadratavvik for interaksjonen år\*stasjon, som er bestemt med 17 frihetsgrader, er vesentlig lavere, mens den skulle være minst like stor som residualvariansen selv om den bare skyldes utvalgusikkerhet mellom replikater. Det tyder på at estimert residualvariens er for stor og ikke representativ for avvik mellom replikater.
2. Siden interaksjonen År\*Stasjon i den første analysen ikke er signifikant, er det gjort en ny analyse på de samme dataene hvor bare hovedeffektene år og stasjon er med som hhv. tilfeldig og fast faktor. Interaksjon år\*stasjon inngår i residualvariens, hvor den er slått sammen med varians mellom replikaterne innen år og stasjon. Residualvariasjonen vil her også omfatte eventuelle langsiktige endringer i forholdet mellom stasjonene. Denne analysen gir residualvariens ca. 0.16 for kombinasjonsmålene (80 % konfidensintervall fra 0.11 til 0.25). I disse to analysene vil estimert variasjon mellom år inkludere både uregelmessige fluktasjoner og jevne tidstrender og den har derfor ikke interesse i sammenheng med analyser av utagskraft.
3. En spesiell analyse er gjort bare på de kombinasjoner av år og tid hvor det er "ekte" replikater tatt i samme måned og år, dvs. Risøyodden 1989 og Croftholmen og Risøyodden 2002, med kombinasjon av stasjon, år og måned som faktor. Her er komponentene 123478/123479-HxCDF og 123678-HxCDF holdt utenfor, fordi de ligger under påvisningsgrense for noen av disse prøvene. Residualvariansen her blir mye mindre enn i de to første analysene slik en vil vente, men pga. stor usikkerhet er 80-% konfidensintervaller fra de ulike analysene overlappende.
4. En kovariansanalyse med log-lineær regresjon mot år er gjort på data fra og med 1991 for de Croftholmen, Klokkartangen og Helgeroa, de tre stasjonene som har mest data. Alle data er med uansett måned, inkludert replikater, med unntak av dobbeltanalyser på samme prøve. Det er lagt inn uavhengig regresjonskoeffisienter for hver stasjon. Kovariansanalysen gir log-lineære regresjonskoeffisienter for tidstrend omkring -0.1, litt varierende mellom de forskjellige komponentene og kombinasjonene. For Croftholmen og Helgeroa, som har best

dekning i tid, blir regresjonskoeffisientene for log(GM3) estimert til hhv. -0.12 og -0.05 , med standardavvik 0.02. Regresjonskoeffisientene tilsvarer en midlere reduksjon av konsentrasjoner på hhv. 12 og 5 % pr. år

Residualvariansen fra denne analysen vil inkludere avvik fra log-lineært forløp av trendene og uregelmessige variasjoner mellom år rundt regresjonen, både felles for alle stasjoner og uavhengig for hver stasjon, og i tillegg varians mellom replikater fra samme stasjon og år. Et forsøk med å trekke inn i analysen log(fett-%) som forklaringsvariabel i regresjonsmodellen gir ikke noen antydning til signifikant sammenheng. En sekundær toveis variansanalyse er gjort på residualene fra denne analysen, med år og stasjon som kategoriske, ”tilfeldige” faktor. Hensikten er å se om det er signifikant uregelmessig varians mellom år som er felles for stasjonene. Det er ikke tegn til det, dvs. at fluktusjon mellom år som gjennomsnitt over stasjoner forklares av interaksjonen stasjon\*år. Dersom det er en slik variasjon, er variansen på logskala anagelig mindre enn 0.025 (standardavvik 0.16). Dette estimatet får en ved å se bort fra at tilfeldig variasjon for interaksjon år\*sted også bidrar til restvariens mellom år.

Resultatene for residualvariens fra de forskjellige analysene som det er valgt å benytte er sammenfattet i **Tabell 7**. Variansanalyse 2 gir residualvariens 0.15 til 0.35 for enkeltkomponentene og ca. 0.16 for kombinasjonene av 3 komponenter. Disse variansestimatene er bestemt med 21 frihetsgrader, og er derfor relativt presise. Variansanalyse nr. 3 på bare ekte replikater gir lavere residualvariens, ca. 0.04 til 0.11 for enkeltkomponenter, og ca. 0.05 for kombinasjonene av 3 komponenter. Kovariansanalysen gir estimater for residualvariens som er noe lavere enn variansanalyse 2.

**Tabell 7.** Estimater og 80-% konfidensintervaller for residualvariens i log(dioksin) for tre ulike analyser på helt eller delvis sammentfallende datamateriale for blåskjell fra Grenland. Residualvariansen omfatter forskjellig bidrag i de ulike analysene, som angitt i beskrivelsen ovenfor

Analysert variabel	Variansanalyse nr. 2 21 frihetsgrader		Variansanalyse nr. 3 5 frihetsgrader, ekte replikater, mellom		Kovariansanalyse med regresjon mot år, residualvariens med	
	Estimat	80 % konf.intervall	Estimat	80 % konf.intervall	Estimat	80 % konf.intervall
log(2378-TCDF)	0.18	0.12 0.28	0.061	0.033 0.19	0.121	0.09 0.18
log(12378/12348-PeCDF)	0.20	0.14 0.32	0.114	0.061 0.35	0.174	0.13 0.25
log(23478-PeCDF)	0.15	0.10 0.23	0.038	0.021 0.12	0.100	0.07 0.15
log(123478/123479-HxCDF)	0.34	0.24 0.53			0.284	0.21 0.41
log(123678-HxCDF)	0.30	0.22 0.48			0.244	0.18 0.35
log(234678-HxCDF)	0.23	0.16 0.37	0.045	0.024 0.14	0.199	0.15 0.29
log(TE3)	0.15	0.11 0.24	0.045	0.024 0.14	0.099	0.07 0.14
log(Sum3)	0.16	0.11 0.26	0.057	0.031 0.18	0.107	0.08 0.16
log(GM3)	0.16	0.11 0.25	0.058	0.031 0.18	0.104	0.08 0.15

5.2 Data fra Kristiansandsfjorden

5.2.1 Beskrivelse av data

Dioksiner er analysert på 15 forskjellige prøver fra Kristiansandsfjorden; de fleste er fra Hanneviksbukta og Vesterhavna. **Tabell 8** viser en oversikt over prøvene fordelt på stasjoner og år. Tabellen viser også hvordan stasjonene er gruppert innenfor delområder; for prøvene fra overvåkingen av Kristiansandsfjorden er dette definert i overvåkingrapportene (Knutzen et al. 1994a, 1998b)

**Tabell 8.** Oversikt over dioksindata for blåskjell fra Kristiansandsfjorden med fordeling på stasjoner og år. Tallet for hver kombinasjon av stasjon og år angir antall analyser.

Delområde	Stasjonsnavn	1986	1988	1990	1992	1995	1996	2002
A.	Myrodden	1		1				
	Fiskåtangen					+1	+1	
	Svensholmen							+1
B.	Silokaia		1	1	2*		1	
	Odderøy vest					+1	+1	+1
D.	Thorsteinsnes		1					
E.	Olafsholman		1					
G.	Ny Hellesund		1					

+n: Prøver fra JAMP-programmet  
\*: Parallell-analyse av samme prøve av NILU og IFFH.

De fleste prøvene er tatt i september-november. Unntaket er 1990, da prøvene ble tatt i mars, dvs. at de tilhører sesongen 1989/90.

Prøvepunktene i område A og B er ikke de i overvåkingen av Kristiansandsfjorden og i JAMP-programmet. Ved Kristiansandsfjordovervåkingen ble det i følge beskrivelse og kart i rapportene tatt prøver på spissen av Myrodden og ved Silokaia på Odderøya. Etter nærmere gjennomgang av JAMP-databasen i 2005 (Norman Green, pers. medd.) ble det i JAMP-programmet for 1995 og 1996 tatt prøver ved Fiskåtangen ca. 300 m sørvest for spissen av Myrodden og ved Odderøya fyr ca. 500 m sør for sørenden av Silokaia. I 2002 ble det ikke funnet blåskjell på Fiskåtangen, og i stedet ble det tatt blåskjellprøve på Svensholmen som ligger ute i fjorden ca. 800 m sørøst for spissen av Myrodden. JAMP-stasjon Odderøy vest ligger ca 400 m sør for sørenden av Silokaia.

Fra Kristiansandsfjorden er det bare for de indre områdene (A, B) at data kan gi et visst bilde av variasjonen fra år til år på samme sted. JAMP-materialet er avgjørende her, men pga. at posisjonene har variert, er det likevel begrenset hva en kan si om dette. For de ytre områdene (D, E, G) ble det analysert på dioksin bare i 1988, så data herfra kan ikke utnyttes for å si noe om variabilitet over tid. Ved å anta samme relative variabilitet som for andre områder kan en likevel prøve å si noe om evenen til å påvise endringer ved fornyet prøvetaking her.

For blåskjell er tørrvekts-% eller fett-% tilgjengelig som biologisk karakteristikk for et flertall av prøvene. En tidligere statistisk analyse av TE for blåskjell i Grenland ga ingen signifikant sammenheng mellom konsentrasjoner på våtvekt og fett-% (Knutzen et al. 2001, Bjerkeng og Ruus 2002). Fra Kristiansandsfjorden er det for få prøver til å etterprøve dette, og variasjon i tørrvekts-% og fett-% er tildels korrelert med geografiske forskjeller på samme måte som dioksinkonsentrasjonene. Prøvene fra de ytre områdene D, E og G har vesentlig lavere dioksinkonsentrasjoner på våtvektsbasis enn prøvene fra område B på omtrent samme tid, men også lavere tørrvekts-% og høyere fett-% enn de fleste prøvene fra de indre områdene A og B. Det er alt for lite data til å skille fra hverandre betydningen av forskjeller i biologisk karakteristikk og forskjeller i dioksinkonsentrasjoner i vann og føde for dioksinkonsentrasjonene i blåskjell i Kristiansandsfjorden. Analysen av dioksin-konsentrasjoner i blåskjell baseres derfor direkte på våtvektskonsentrasjoner.

## 5.2.2 Statistiske analyser

Siden det stort sett ikke er tatt flere prøver på samme tid og sted, er det ikke grunnlag i data fra Kristiansandsfjorden til å estimere variasjon mellom replikater tatt på samme tid og sted. Det betyr at det ikke er grunnlag for å analysere betydningen av å øke prøvematerialet bak hver analyse, for eksempel ta blandprøver på tvers av flere gjentatte prøvetakinger. Derimot er det mulig å få estimater for uregelmessig variasjon over tid uavhengig for hver stasjon, ved å se på interaksjonen tid\*sted i en variansanalyse. Interaksjonen vil være en kombinasjon av utvalgsusikkerhet innenfor tid og sted og variasjon i gjennomsnittsnivå mellom ulike kombinasjoner av tid og sted. Det kan også gi et visst grunnlag for å vurdere om variasjonen mellom prøver fra år til år er korrelert mellom stasjoner eller områder eller om den er preget av uregelmessige utslag som er forskjellig fra stasjon til stasjon.

Det kan gjøres to variansanalyser, på delvis overlappende materiale, for å estimere residualvarians:

1. En kan sammenligne Fiskåangen/Svensholmen med Odderøy vest for årene 1995, 1996 og 2002, dvs. bruke prøvene i skravert del av **Tabell 8**. Det gir grunnlag for å estimere interaksjonen sted\*år med 2 frihetsgrader.
2. En kan sammenligne data fra område A med data fra område B for årene 1990, 1995, 1996 og 2002, dvs. bruke de prøvene som står i ruter med dobbel ramme i **Tabell 8**. Det omfatter de samme data som i analyse 1, men med noen prøver i tillegg. Det gir grunnlag for å estimere interaksjon område\*år med 4 frihetsgrader. Prøvene fra Silokaia og Odderøy vest fra 1996 behandles da som replikater fra samme område og år.

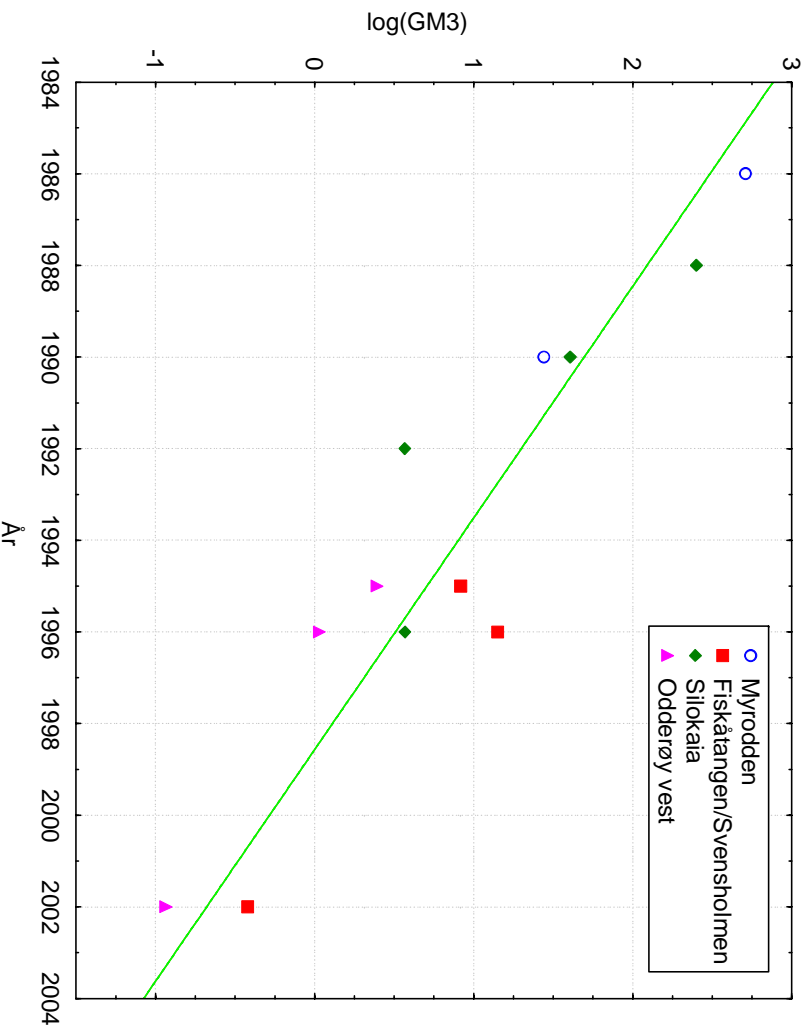
Begge analysene vil bygge på en variansmodell med to hovedeffekter sted og år:

$$\log(C) = \alpha_{\text{sted}} + \beta_{\text{år}} + \varepsilon_{\text{sted,år}} \quad (1)$$

Analysene er gjort både på de aktuelle enkelt-komponentene og på de ulike kombinasjonsmålene av dioksiner. I hver analyse estimeres variansen for det siste leddet, interaksjonsleddet  $\varepsilon_{\text{sted,år}}$ , som et residualledd som de andre effektene testes mot. Residual-leddet omfatter både statistisk utvalgsusikkerhet (mellom replikater) og varians pga. at gjennomsnittskonsentrasjonen på forskjellige steder varierer uavhengig av hverandre. Mangelen på replikater gjør at det ikke er grunnlag for å skille de to bidragene fra hverandre.

I disse analysene estimeres også en varians for felles variasjon mellom år, men den er i hovedsak et uttrykk for en nedadgående tidstrend, dvs. det signalet en vil ønske å påvise, og sier ikke noe om fluktasjoner mellom år. Det har derfor ikke interesse som grunnlag for å vurdere hvordan overvåkingen bør legges opp for å minimere de uregelmessige variasjonene.

Variasjonen over tid for hver stasjon er i stor grad en lineær variasjon på logskala. **Figur 4** viser den kombinerte konsentrasjonen  $\log(\text{GM3})$  mot observasjonsår, med en tydelig log-lineær nedgang over tid for områdene A og B (Myrødden, Fiskåangen, Silokaia og Odderøy vest), med nok så små forskjeller mellom stasjoner eller områder.



**Figur 4.** Variasjon mot kalenderår for lnGM3) i blåskjell fra Kristiansandsfjorden. Regresjonslinje fra kovariansanalyse tegnet inn, med gjennomsnittlig nivå for område A og B.

Ut fra dette er det gjort en kovariansanalyse med forskjell mellom steder som kategorisk faktor og felles log-lineær regresjon mot år for alle data fra område A og B samlet. Det er ikke grunnlag for å estimere forskjeller mellom områder i regresjonskoeffisient mot år. Den modellen som analyseres er:

$$\log(C) = a_{sted} + b \cdot \text{År} + \varepsilon_{sted, \text{år}} \quad (2)$$

Residualvariansen her omfatter både utvalgsusikkerhet, uregelmessig variasjon i tid som er felles for alle områdene, og svinginger som skjer uavhengig for hver stasjon, i tillegg vil også avvik fra log-linær trend og forskjeller mellom stasjoner i størrelsen på tidstrendene inngå i residualvariansen i denne modellen. Det er derfor grunn til å tro at denne variansen er større enn den støykomponenten vi er ute etter. Disse elementene kunne vært trukket ut som egne effekter i en mer kompleks modell, men da ville antall frihetsgrader minske sterkt. Den enklere kovariansanalysen gir et mer presist estimat selv om det vil tenderer mot å ligge for høyt i forhold til reell residualvarians. Det gir antagelig også et ganske realistisk resultat i forhold til en analyse av tidstrender på nye data, siden en også da vil ha problemer med å skille ut slike elementer fra residualvariansen.

Tabell 9 viser residualvarians med estimat og konfidensintervall for enkeltkomponenter og kombinasjoner av 3 komponenter fra de ulike analysene. Den første variansanalysen estimerer varianser for interaksjonsleddet som spenner fra 0.022 til 0.12 på logskala for de forskjellige dioksin-komponentene sett enkeltvis og 0.06 for kombinasjonsmålene. Disse estimatene har bare to frihetsgrader, og 80-% konfidensområder går derfor fra 0.43 til 9.5 ganger estimert varians (Vedlegg F). Den andre variansanalysen, som omfatter flere data, estimerer varianser for interaksjonsleddet område\*år fra 0.05 til 0.49 for de forskjellige enkeltkomponentene og ca. 0.11 for kombinasjonsmålene. Med 4



frithetsgrader går 80-% konfidensområdene fra 0.51 til 3.76 ganger estimert varians. I variansanalyse 2 vil evt. permanent forskjell mellom stasjonene innenfor et område inngå i den estimerte interaksjonen, fordi stasjonene innen et område dekker forskjellige år, så det er rimelig at variansestimaterne fra analyse 2 er større enn fra analyse 1. Imidlertid er usikkerheten i estimatene fra variansanalyse 1 stor, så det kan være tilfeldig. Kovariansanalysen finner residualvarianser som ikke er så mye større enn Variansanalyse nr. 2, men med 9 frithetsgrader og derfor bedre presisjon; 80% konfidensintervallet går fra 0.61 til 2.16 ganger estimert varians. Ingen av analysene gir signifikant forskjell mellom områdene sammenlignet med interaksjonsleddet. Det er også gjort en totrimsanalyse hvor residualene fra kovariansanalysen med log-lineær regresjon mot år etterpå analyseres i en variansanalyse med år som tilfeldig faktor; dette viser ingen signifikant varianskomponent mellom år som gjennomsnitt over områder.

Konfidensintervallene fra alle de tre analysene er i stor grad overlappende. Det må bemerkes at de ulike resultatene i tabellen er forskjellige analyser på helt eller delvis overlappende materiale for de samme prøvene, og ikke uavhengig av hverandre. Avvik fra riktig verdi vil derfor delvis også være koblet til hverandre, slik at f.eks. forskjellen mellom de tre analysene kan være mer reell enn konfidensintervallene gir inntrykk av.

**Tabell 9.** Estimater og 80-% konfidensintervaller for residualvarians for tre ulike analyser på helt eller delvis sammenfallende datamateriale for blåskjell fra Kristiansandsfjorden.

Analysert variabel	Variansanalyse nr. 1 residualvarians med 2 frithetsgrader		Variansanalyse nr. 2 residualvarians med 4 frithetsgrader		Kovariansanalyse med regresjon mot år, residualvarians med 9 frithetsgrader	
	Estimat	80 % konf.intervall	Estimat	80 % konf.intervall	Estimat	80 % konf.intervall
log(2378-TCDF)	<b>0.06</b>	0.026 0.58	<b>0.05</b>	0.027 0.198	<b>0.10</b>	0.060 0.211
log(12378/12348-PeCDF)	<b>0.02</b>	0.011 0.23	<b>0.12</b>	0.059 0.434	<b>0.26</b>	0.159 0.561
log(23478-PeCDF)	<b>0.12</b>	0.053 1.16	<b>0.23</b>	0.118 0.867	<b>0.17</b>	0.102 0.360
log(123478/123479-HxCDF)	<b>0.02</b>	0.009 0.21	<b>0.37</b>	0.189 1.380	<b>0.39</b>	0.237 0.836
log(123678-HxCDF)	<b>0.07</b>	0.030 0.66	<b>0.46</b>	0.236 1.725	<b>0.46</b>	0.280 0.986
log(234678-HxCDF)	<b>0.06</b>	0.027 0.60	<b>0.49</b>	0.254 1.858	<b>0.47</b>	0.290 1.021
log(TE3)	<b>0.08</b>	0.033 0.72	<b>0.11</b>	0.054 0.398	<b>0.12</b>	0.071 0.250
log(SUM3)	<b>0.06</b>	0.025 0.55	<b>0.07</b>	0.035 0.253	<b>0.11</b>	0.067 0.235
log(GM3)	<b>0.06</b>	0.026 0.57	<b>0.11</b>	0.057 0.420	<b>0.14</b>	0.088 0.309

Kovariansanalysen gir i tillegg til residualvarians også regresjonskoeffisienter for reduksjon med tid (b) som varierer fra -0.18 til -0.23; for de kombinerte dioksimålene er  $b = -0.20$ . Regresjonskoeffisientene er klart signifikante, men for enkelt-komponentene er de ikke bestemt med høy presisjon; standardavvik for regresjonskoeffisientene er mellom 0.05 og 0.1. For de kombinerte målene TE3, Mean3 og GM3 har regresjonskoeffisienten bedre presisjon med et standardavvik ca. 0.02, dvs. at vi med ca. 80 % konfidens kan anslå regresjonskoeffisienten til å ligge mellom 0.15 og 0.22. Det må bemerkes at hovedhensikten med kovariansanalysen er å få et alternativt estimat for residualvariansen med en litt stivere modell som vil tendere i retning av å overestimere den. Siden nedgangen skjer samtidig med store utslippsreduksjoner (se kapittel 1) er det ikke grunn til å tro at det skal være noen jevn nedgang over tid slik kovariansmodellen forutsetter, og reduksjonratene representerer neppe forventet reduksjon i en perioden etter utslippsreduksjoner. De lave verdiene i blåskjell for 2002 faller sammen med spesielt lave rapporterte utslipp fra Falconbridge, men konsentrasjonene er totalt sett redusert mindre enn utslippene når en ser på hele perioden 1988 til 2002, så det kan også være en forsinket effekt av utslippsreduksjonen omkring 1992. Videre over-

våking vil avklare om de lave verdiene i 2002 er en midlertidig følge av lave utslipp eller en mer permanent forsinket nedgang etter total utslippsreduksjon.

Blåskjelldataene fra Kristiansandsfjorden er for det meste fra september og oktober; bare to prøver er tatt i mars og april. Det er ikke mulig ut fra så lite data å se om måned har noe å si for nivåene.

Antall skjell i hver prøve varierer fra ca. 60 til 120 der hvor antallet er oppgitt. Generelt vil økende antall skjell redusere utslaget av variasjon mellom enkeltskjell for gjennomsnittet i prøven, men det er ikke tegn til det i dette materialet, det måtte også ganske kraftige forskjeller til for å se det ut fra så få data. Hvis det er stor variasjon i miljøgiftkonsentrasjon og bløtdelsvekt mellom individer, vil det ikke bli så effektiv utjevning av konsentrasjoner ved å øke individantallet i prøven som for blanding av ikke deler fra hvert individ og symmetrisk fordeling av konsentrasjoner på individnivå (Vedlegg H).

## 5.3 Sammenfattende konklusjon for blåskjelldata

De tre analysene på Grenlandmaterialet med ulike modeller gir alle overlappende konfidensintervaller for residualvariens (**Tabell 7**), og det er derfor ikke noe sikkert grunnlag for å skille fra hverandre ulike varianskomponenter. Forskjellene i variansestimater fra variansanalyse 3 og kovariansanalysen kan likevel brukes til å antydes en oppdeling i ulike varianskomponenter. Det riktigste estimatet for varians for avvik fra langsiktig trend i enkeltp prøver er antagelig det som kommer fra kovariansanalysen, dvs. 0.11 for logtransformerte verdier av kombinasjonsmålene TE3, Sum3 og GM3. Siden kovariansanalysen i liten grad bygger på replikate prøver fra samme sted og tid, vil variansestimater være en sum av varians mellom replikate prøver fra samme stasjon og tid (utvalgusikkerhet) og varians for uregelmessige avvik mellom ulike kombinasjoner av stasjon\*år fra gjennomsnittlige trender og nivåer. Variansanalyse 3 har residualvariens på ca. 0.06 som kan ses som estimat for varians mellom ekte replikate prøver, og den reelle uregelmessige variasjonen over tid og sted kan da settes til differansen som er 0.05, i hovedsak med uavhengige utslag på hver stasjon.

Analysene av data fra Kristiansandsfjorden estimerer bare total varians for interaksjon stasjon\*år, dvs. en sum av varians i over tid og sted og varians mellom replikate prøver, og de tre analysene gir konfidensintervaller for varians som stemmer godt overens innbyrdes og med estimatene fra Grenlandmaterialet (**Tabell 9**). For de tre kombinasjonsmålene blir det mest sannsynlige estimatet for residualvariens ca. 0.12 med 80% konfidensintervall fra 0.07 til 0.3.

Alt i alt virker det rimelig å regne med at uregelmessige flukturasjoner for blandprøver å 50 skjell fra forskjellige kombinasjoner av stasjon\*år ligger i intervallet 0.08 til 0.16 på naturlig logskala, med 0.11 som mest sannsynlig verdi. Dette er da varians for uregelmessige avvik fra gjennomsnittlig, jevn tidsutvikling på de forskjellige stasjonene. Det er ikke funnet sikkert grunnlag for å anta at det er noen vesentlig felles komponent over flere stasjoner i denne uregelmessige variasjonen, eller at varians for interaksjon stasjon\*år er større enn det som kan forklares av utvalgusikkerheten ved prøvetaking. Det kan likevel antydes en oppdeling i en reell varians for interaksjon stasjon\*år på 0.05 og en utvalgusikkerhet, dvs. variasjon mellom replikater fra samme tid og sted med varians 0.07 mellom replikate prøver fra samme sted og tid. Estimater for variansen mellom replikater er basert på 5 frihetsgrader justeres opp fra 0.06 til 0.07 for å kompensere for tendensen til å underestimere ved få frihetsgrader (Vedlegg F). Oppdelingen er som sagt usikker, og det er ikke grunnlag for å si hvor mye av variansbidraget for interaksjon stasjon\*tid som gjelder mellom måneder innenfor år og hvor mye som er mellom år. Dette skyldes at datasettet er uregelmessig, slik at ulike varianskombinasjoner må estimeres på ulikt data-grunnlag. Hvis variasjonene har en felles komponent som gjennomsnitt over stasjoner, ser det ut til at den har varians mindre enn 0.025.

For å minke uregelmessige variasjoner kan en øke antall skjell pr. prøve, men det er begrenset hva en kan oppnå med dette. Hvis en går ut fra en varians 0.06 på naturlig logskala for replikate prøver med 50 skjell, og bruker resultatene i Vedlegg H om egenskaper til blandprøver når konsentrasjonene er lognormalt fordelt på individnivå, kan varians på logskala mellom individuelle skjell anslås til

$\sigma^2=1.55$ , og det vil si at varians for blandprøver med  $N$  individer i følge Vedlegg H blir  $\sigma^2/N^{0.83}$ .

Dersom en da i stedet bruker 200 skjell pr. prøve, skulle replikativariansen på logskala reduseres til ca. 0.02. Samlet varians for blandprøver fra ulike kombinasjoner av stasjon\*år vil da reduseres fra 0.11 til  $0.05+0.02=0.07$ , og det betyr at standardavvik på logskala reduseres fra 0.33 til 0.26.

For data fra Grenlandsområdet estimeres reduksjonstrater på 5 til 12 % pr. år av konsentrasjoner i blåskjell for en 10-årsperiode etter de store utslippreduksjonene. For Kristiansandsfjorden estimeres større reduksjonstrater, men med stor usikkerhet og fra en periode med utslippreduksjoner. Som forsinket effekt de første 10 år etter store utslippreduksjoner kan 10 % reduksjon pr. år av konsentrasjoner anses å være realistisk.

De statistiske analysene av tidstrender for dioksininnhold i blåskjell basert på datamaterialet fra Kristiansandsfjorden og Grenlandsfjordene tilsier:

- Hvis tidsutviklingen i dioksinkonsentrasjoner på ett sted overvåkes ved at en tar blåskjellprøver over en årrekke, med 50 skjell i hver prøve, vil konsentrasjoner for enkelprøvene variere tilfeldig fra år til år rundt en jevn tidstrend. Disse variasjonene har relativt standardavvik ca. 30 til 40 % (variens 0.11 for naturlige logaritmer av konsentrasjoner).
- Omtrent halvparten av variansen ser ut til å være utvalgusikkerhet pga. tilfeldig variasjon mellom skjell fra samme tid og sted. Den andre halvparten skyldes antagelig reelle avvik fra trenden i gjennomsnittet fra år til år. Hvis det analyseres flere prøver fra samme stasjon hvert år, eller hvis antall individer pr. prøve økes, vil virkningen av utvalgusikkerheten avta, mens virkningen av de reelle avvikene vil være den samme.
- Hvis det tas parallelle tidsserier fra flere stasjoner innenfor et område, kan en regne med at disse avvikene fra tidstrendene er uavhengige på hver stasjon. Virkningen på samlet årsgjennomsnitt vil da minske jo flere stasjoner en har. I tillegg kan det være parallelle avvik som er felles for alle stasjoner innenfor et område (estimert variens 0.025 på naturlig logskala). Et slikt variansbidrag vil ikke ha særlig betydning totalt sett hvis antall stasjoner er 6 eller mindre. Dette er basert på data fra Grenland, en analyse av mye færre data fra Kristiansandsfjorden viser ikke signifikante parallelle avvik over flere stasjoner.
- Data fra Grenland viser ca. 10 % reduksjon av blåskjellkonsentrasjoner pr. år over en 10-årsperiode etter store utslippreduksjoner. Det antas at det er realistisk å vente det samme i Kristiansandsfjorden etter de store utslippreduksjonene som har skjedd der siden 1990.

## 6 Statistisk analyse av data for torsk

### 6.1 Data fra Grenland

Data for torsk fra Grenland er for det meste stedfestet bare til fjordområder. **Tabell 10** viser hvordan datamaterialet fordeler seg over tid og område. Lange tidsserier har vi fra Frierfjorden, Breviksfjorden og Langesundsbukta. I tillegg er det noen data fra de siste årene fra Eidangerfjorden. De aller fleste prøvene er blandprøver, stort sett med omkring 20 torsk i hver prøve. I tillegg til de data som er vist i denne oversikten er det også analysert noen torskeleverprøver fra området rundt Jomfruland (Dypbåen 1988, Kragerø 2002, Klokkartangen og Jomfruland 2003). Prøvene fra 1975, 1976, 2002 og 2003 er tatt om høsten (september-desember), og det samme gjelder to av prøvene fra Langesundsbukta 1988 og en prøve fra Frierfjorden 2000, ellers er de fra april eller mai.

**Tabell 10.** Oversikt over dioksindata for torskelever fra Grenland med fordeling på delområder og kalenderår. Tallet for hver kombinasjon av delområde og år angir antall analyser.

Delområde	1975	'76	'87	'88	'91	'92	'93	'94	'95	'96	'97	'98	'99	2000	'01	'02	'03
Frierfjorden	1	1	6		1	1	2	1	1	1	1	1	1	8	1	1	3
Breviksfjorden				1	1	1	1	1	1	1	1	1	1	1	1		1
Eidangerfjorden														8		1	
Langesundsbukta				4	1	1	1	1	1	1	1	1	1	1	1	1	1
Jomfrulandområdet				1													

Det er noen utfeller av flere prøver fra samme tid og sted, bl.a. fra Frierfjorden i 1987 med 6 individuelt analyserte torsk, og fra april og juni 2000 med 8 prøver å 5 torsk fra hvert av områdene Eidangerfjorden og Langesundsfjorden, samlet inn i prosjektet "Dioksiner i Grenland" (DIG). **Tabell 11** viser en nærmere oversikt over tilfellene av replikate prøver. For individuelt analyserte torsk fra 1987 er lengde og vekt er registrert, men ikke fett-%. For prøvene fra DIG-prosjektet er det er registrert lengde, totalvekt og levervekt for de fisk som inngår i prøven, og fett-% i blandprøven av lever. For de fleste prøvene er gjennomsnittlig lengde, vekt og fett-% oppgitt.

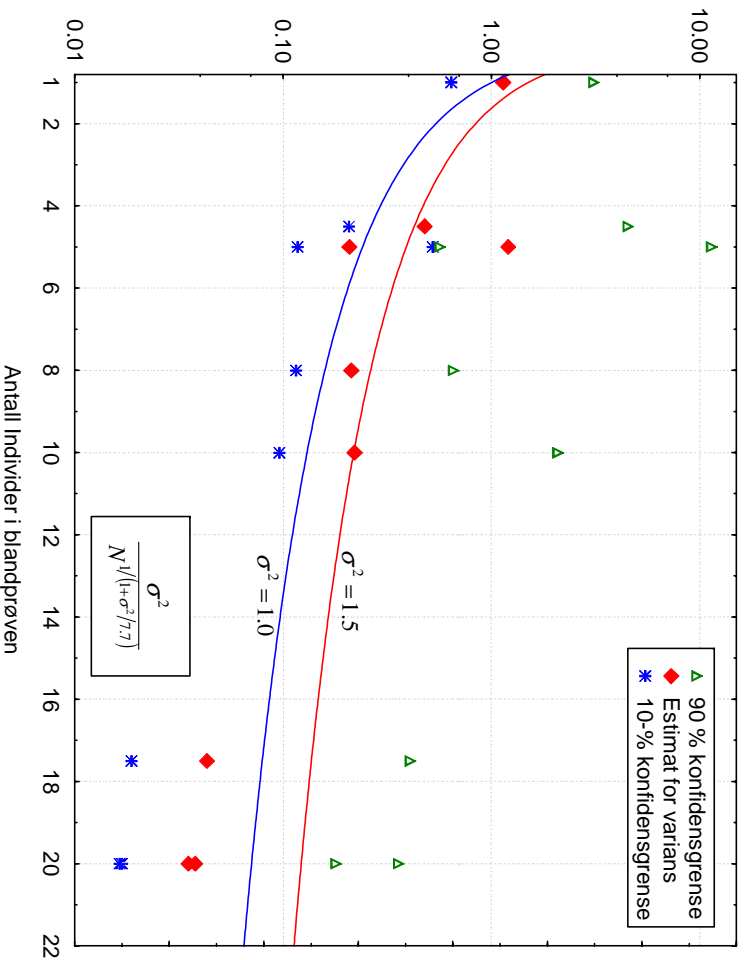
**Tabell 11.** Oversikt over sett av replikat-prøver i datamaterialet for torskelever fra Grenland.

År	Måned	Område	Gjennomsnittlig antall individer pr. prøve	Antall replikate prøver
1987	5	Frierfjorden	1	6
1988	10	Såstein	10	2
2000	4	Eidangerfjorden	5	6
2000	4	Frierfjorden	8*	5
2000	6	Eidangerfjorden	5	2
2000	6	Frierfjorden	4,5	2
2003	10	Frierfjorden	20	3
2003	10	Klokkartangen	20	2
2002	11	Kragerø	17,5	2

\*: 4 prøver å 5 fisk, 1 prøve å 20 fisk

### 6.1.1 Analyse av replikate prøver fra samme område og år

Som **Tabell 11** viser, varierer antall individer pr. blandprøve mellom de forskjellige settene av replikate prøver. Det betyr at en ikke uten videre kan betrakte dem under ett for å estimere varians mellom replikater uten å ta det i betraktning. Det er funnet ved numerisk simulering at hvis konsentrasjonene på individnivå er log-normalfordelt med varians  $\sigma^2$  på logskala vil logtransformerte konsentrasjoner i blandprøvene vil ha varians  $\sigma^2 / N^{1/(1+\sigma^2/7.7)}$  (se Vedlegg H). **Figur 5** viser varians på logtransformerte verdier av kombinasjonsmålet GM3 for de forskjellige settene av replikate prøver i **Tabell 11** som funksjon av gjennomsnittlig antall individer i prøven. Øvre og nedre grense for 80-% konfidensintervaller ut fra Vedlegg F er også markert. Variansen reduseres med økende antall individer i prøven, og intervallene er konsistent med et forløp som beskrevet, med varians mellom logtransformerte konsentrasjoner i individer på 1.0 til 1.5.



**Figur 5.** Estimator og 80-% konfidensintervaller for varians i log(GM3) mellom replikat-prøver som funksjon av antall individer i prøvene. Teoretisk forløp ved gitt varians mellom individer med lognormal og normal fordeling tilpasset visuelt.

Det er også gjort en litt nærmere analyse av de to største datasettene av replikater, dvs. de 6 individuelt analyserte fiskene fra 1987 og blandprøver å 5 fisk fra DIG-prosjektet **Tabell 12** viser resultatet av separate variansanalyser for de to datasettene for log(GM3). Variansen i log(GM3) mellom individene fra 1987 estimeres til 1.14. Blandprøvene å 5 fisk fra DIG-prosjektet i 2000 er analysert som 4 sett av replikate prøver: 6 prøver fra Eidangerfjorden og 5 fisk fra Frierfjorden i april, og 2 fra hvert område i juni, det gir estimat for residualvariens av log(GM3) på 0.31. I tabellen er hvert resultat brukt til å anslå varians for andre prøvestørrelser. De tallene som står med fet skrift er de direkte estimatene ut fra de to datasettene. Verdier i kursiv er beregnet ut fra beskrivelsen i Vedlegg B av egenskapene i blandprøver som representerer et aritmetisk gjennomsnitt av lognormalt fordelte konsentrasjoner i enkeltfisk. For enkeltfisk fra 1987 er varians for blandprøver beregnet direkte ut fra varians for

enkeltfisk. For blandprøvene fra 2000 er varians for enkeltfisk funnet ved å iterere til en verdi som stemmer med estimert varians mellom blandprøvene. Nedre og øvre konfidensgrenser er beregnet for enkeltfisk, og så regnet om til blandprøver ut fra beregnede reduksjonsfaktorer i tabellen. Til slutt er det beregnet kombinert estimat med 80-% konfidensintervall.

Omregningen viser at de to datasettene gir estimater som stemmer godt med hverandre. En vektet kombinasjon ut fra de to datasettene gir variansestimater for  $\log(\text{GM3})$  i enkeltfisk på ca. 1.2 med 14 frihetsgrader og 80-% konfidensintervall fra 0.8 til 2.15. Det stemmer rimelig bra med det en finner for innholdet av HCB, OCS og DCB i torskelever fra Frierfjorden ut fra et mye større datamateriale for enkeltfisk: varians mellom individer fanget på samme tid og sted er hhv. 1.1, 0.85 og 0.75. Hvis vi baserer oss på estimatet på 1.2 for dioksiner, vil variansen mellom replikate prøver reduseres med antall individer pr. prøve ( $N$ ) etter en funksjon  $\sigma^2/N^{0.87}$ .

$N$  blandprøver av 5 og 20 fisk gi reduksjon av variansen mellom replikate prøver med faktorer hhv. ca. 4.0 og 13.4 (nederste linje i tabellen). Residualvariansen for blandprøver å 20 fisk anslås til 0.09 med 14 frihetsgrader; det representer utvalgsusikkerhet knyttet til variasjon mellom individer fanget på samme tid og sted.

Analysen av DIG-dataene gir også et estimat for varians mellom forventningsverdier for forskjellige repikat (sted\*tid) som for  $\log(\text{GM3})$  blir 0.1. Dette estimatet er imidlertid ikke signifikant større enn 0, dvs. at forskjell i observert gjennomsnitt for ulike sett av replikate prøver ligger innenfor det som kan forklares som et tilfeldig utslag av utvalgsusikkerheten mellom replikate prøver innenfor sted\*tid.

**Tabell 12.** Estimater for varians i logtransformerte dioksinkonsentrasjoner i replikate blandprøver av torskelever i Frierfjorden som funksjon av antall individer.

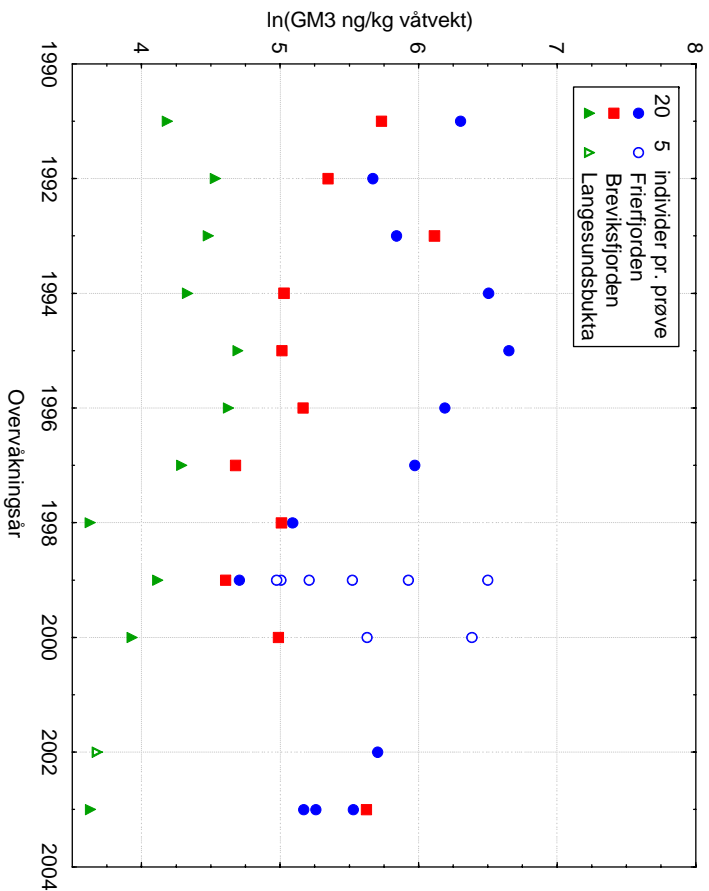
		Type estimat	Varians for logtrans- formert konsent- rasjon i enkeltfisk	Reduksjonsfaktor for varians av blandprøver (=effektivt antall individer)	Varians for log- transformert konsentrasjon i blandprøver		
Enkeltfisk i 1987	5	Nedre 80% konf.grense	<b>0.62</b>	4.4	16.0	<b>0.14</b>	0.04
		Estimert verdi	<b>1.14</b>	4.1	13.6	0.28	0.08
Blandprøver å 5 fisk fra 2000	9	Øvre 80 % konf.grense	<b>3.54</b>	3.0	7.8	1.18	0.45
		Nedre 80% konf.grense	0.75	4.3	15.3	<b>0.17</b>	0.05
		Estimert verdi	1.23	4.0	13.2	<b>0.31</b>	0.09
Kombinerte estimater	14	Øvre 80 % konf.grense	2.66	3.3	9.3	<b>0.80</b>	0.29
		Nedre 80% konf.grense	0.80	4.3	15.1	0.19	0.05
		Estimert verdi	1.20	4.0	13.4	0.30	0.09
		Øvre 80 % konf.grense	2.15	3.5	10.4	0.61	0.21

Verdier i fet skrift er direkte estimert av data

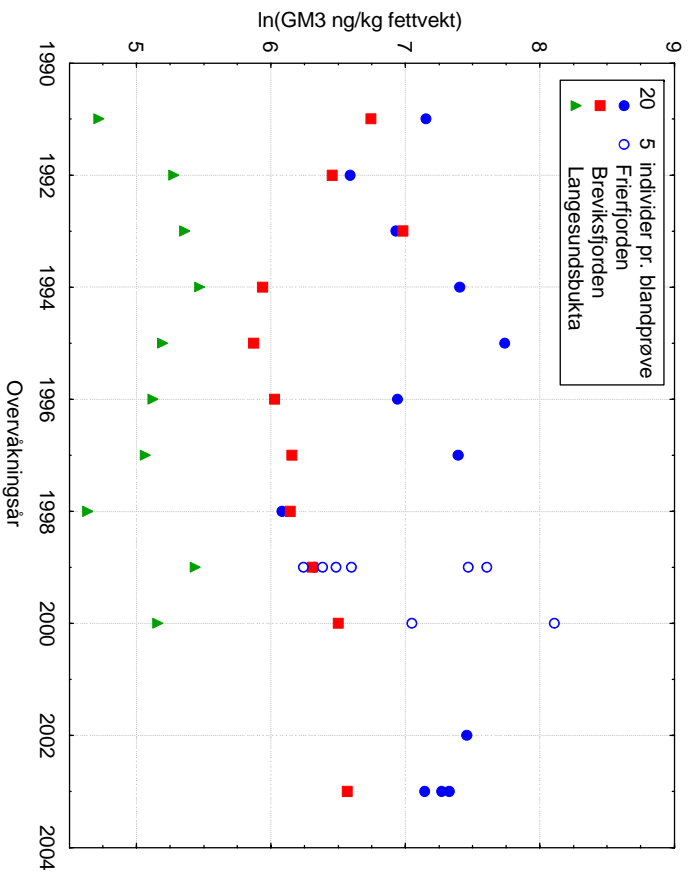
Verdier i kursiv er beregnet av de som er estimert av data

6.1.2    Analyse av stasjoner med lengre tidsserier

Som for blåskjell er det naturlig å se på data fra og med 1991, etter siste utslippsreduksjon, for å estimere varianskomponenter til bruk i planlegging av fremtidig overvåking. **Figur 6** og **Figur 7** viser tidsplott av  $\log(\text{GM3})$  på hhv. våtvektsbasis og fettvektsbasis. De fleste punktene bygger på blandprøver av 20 fisk; replikatenes fra DIG-prosjektet med 5 fisk pr. prøve er markert med åpne symboler.



**Figur 6.** Konsentrasjon mot overvåkningsår for ln(GM3) i torsk fra Grenland, på våtvektsbasis. Prøver fra små blandprøver er vist med åpne symboler



**Figur 7.** Konsentrasjon mot overvåkningsår for ln(GM3) i torsk fra Grenland, på fettvektsbasis. Prøver fra små blandprøver (5 fisk) er vist med åpne symboler. Verdier fra Langesundbukta i 2002 og 2003 mangler fordi fettvekt ikke er oppgitt.

Det visuelle inntrykket av data på våtvektsbasis er en klar generell tendens til reduksjon over perioden sett som helhet, men med så store svingninger frem og tilbake fra år til år at det er vanskelig å se om det pågår en fortsatt reduksjon i siste del av perioden. Når konsentrasjonene er regnet om til fettbasis, er det ingen tegn til nedgang over tid på noen stasjoner. Det henger sammen med at fett-% viser en synkende tendens over tid i prøvene fra Friefjorden og til dels også i Breviksfjorden. Hva som gir det riktige bilde av langsiktig utvikling er usikkert, men i denne sammenhengen er det fluktuasjonene som er mest interessante. Det er valgt å se på våtvektskonsentrasjonene, og data er analysert på to måter:

- En variansanalyse hvor delområde og år brukes som kategoriske faktorer, og med interaksjonsleddet område\*år som residualvarians, slått sammen med variasjon mellom replikater innen kombinasjoner av område og år for de få tilfeller hvor det forekommer. Varians for hovedeffekt år inkluderer tidstrender og har ikke interesse for vurdering av utsagnskraft.
- En kovariansanalyse med tilpasning av log-lineær regresjon mot år, med område som kategorisk faktor. Det er forsøkt både med antatt felles tidstrend for alle områdene, dvs. at det tilpasses en felles regresjonskoeffisient mot år, og med interaksjon område\*år slik at regresjonskoeffisienten bestemmes uavhengig for hvert område.
- En sekundær variansanalyse på residualene fra kovariansanalysen med år og delområde som tilfældige kategoriske faktorer, og med interaksjon inkludert i residualvariansen er gjort for å prøve å skille ut variasjon fra år til år som er felles over stasjonene fra uavhengig variasjon for ulike stasjoner, dvs. interaksjon stasjon\*år.

Observasjoner er vektet ut fra antall fisk ut fra resultatene i kapittel 6.1.1, med vekt=1 for prøver av 20 fisk [vektingsfaktor =  $(N/20)^{1/(1+2/7)}$  ]. Antall frihetsgrader er bestemt ut fra de nedjusterte vektene.

Hvis replikatprøvene fra år 2000 kuttet, blir resultatene ikke særlig forandret.

**Tabell 13** viser estimerte residualvarianser fra de to første analysene. I variansanalysen er det interaksjonsleddet område\*år som utgjør den dominerende delen av residualleddet, og dette omfatter også forskjeller i tidsutvikling mellom delområdene. I kovariansanalysen, som antar log-lineært tidsforløp, er avvikene fra lineært forløp også inkludert i residualleddet. De to modellene gir svært like residualvarianser, og også med lite variasjon mellom de forskjellige enkeltkomponentene, stort sett i området 0.15 til 0.35. For de kombinerte dioksinmålene TE3, Sum3 og GM3 er estimatene for residualvarians 0.16 til 0.18. Fordi data har mange frihetsgrader, er estimatene relativt presise, med 80 % konfidensintervaller innenfor  $\pm 20$  til 30 %. At de stivere regresjonsmodellene ikke gir høyere residualvarians enn ANOVA-modellene tyder på at det ikke er særlig samvariasjon mellom fluktuasjonene fra år til år for de tre områdene. Kovariansanalyse med felles tidstrend gir regresjonskoeffisient -0.08 for log(GM3), det vil si en reduksjon av konsentrasjoner på ca. 8 % pr. år. Det gir ingen bedre tilpasning å innføre uavhengige log-lineære regresjonskoeffisienter for de tre områdene. Den sekundære variansanalysen på residualene fra kovariansanalysen med år og delområde som tilfældig kategorisk faktor gir ingen signifikante fluktuasjoner som er felles for stasjonene, hvis det er noen slik variasjon er den antagelig mye mindre enn 0.02.

Konklusjonen blir at det beste estimatet for varians mellom enkeltp prøver basert på 20 fisk er ca. 0.16, av dette kan det anslås at 0.09 er utvalgsusikkerhet knyttet til variasjon mellom individer som estimert foran, mens varians for fluktuasjoner over tid i gjennomsnittskonsentrasjoner er 0.07. Fluktuasjonene ser i hovedsak ut til å være uavhengig for hver stasjon.



**Tabell 13.** Estimator og 80-% konfidensintervaller for residualvarians i log(GM3 våtvektsbasis) for to ulike analyser på samme datamateriale for torskelever fra Grenland.

Analysert variabel	Variansanalyse residualvarians med ca. 31 frihetsgrader			Kovariansanalyse med felles tidstrend residualvarians med ca. 37 frihetsgrader		
	Estimat	80 % konf.intervall		Estimat	80 % konf.intervall	
log(2378-TCDF)	0.19	0.14	0.28	0.17	0.14	0.25
log(12378/12348-PeCDF)	0.19	0.14	0.28	0.20	0.16	0.29
log(23478-PeCDF)	0.27	0.20	0.40	0.25	0.19	0.35
log(123478/123479-HxCDF)	0.22	0.17	0.34	0.24	0.19	0.35
log(123678-HxCDF)	0.21	0.16	0.31	0.22	0.17	0.31
log(234678-HxCDF)	0.28	0.21	0.42	0.35	0.27	0.50
log(TE3)	0.18	0.14	0.27	0.16	0.13	0.23
log(Sum3)	0.17	0.13	0.25	0.16	0.13	0.23
log(GM3)	0.17	0.13	0.26	0.16	0.13	0.23

6.1.3 Sammenhenger med biologisk karakteristikk eller andre miljøgifter

Biologisk karakteristikk

Både for de utvalgte settene av replikate prøver analysert i kapittel 6.1.1, og langtidsseriene som er analysert i kapittel 6.1.2, er det undersøkt om det å ta med biologiske kovariater i modellen (lengde, vekt, levervekt eller fett-%) reduserer variansen.

For de 6 individuelt analyserte torsk fra Frierfjorden i mai 1987 viser dioksinkonsentrasjonene ingen sammenheng med hverken lengde eller vekt. Materialet fra 2000 gir heller ikke særlig signifikant sammenheng med lengde eller fett-% som kovariater, bare for noen av enkelt-komponentene estimeres signifikant sammenheng med fiskens lengde, men som en reduksjon av konsentrasjonene med økende lengde, og det virker urimelig. Materialet er antagelig for lite til å gi noe pålitelig bilde av eventuelle koblinger til biologisk karakteristikk.

For langtidsseriene fra de tre hovedstasjonene er det brukt en generell regresjonsmodell (GRM) som en utvidelse av kovariansanalysen som ble beskrevet i kapittel 6.1.2, hvor også de biologiske karakteristikkene lengde<sup>1</sup> og fett-% er brukt som kovariater. Som i andre analyser er både konsentrasjonene og de biologiske kovariatene logtransformert. Disse analysene gir signifikant sammenheng med biologiske kovariater for noen av enkeltkomponentene, men litt sprikende. Problemet her er at kovariaterne til dels også har en langsiktig endring over tid. Lengde har økende tendens med tid i prøvene fra Frierfjorden, og avagende i de to andre områdene, og fett-% viser fallende tendens i Frierfjorden og til dels også i Breviksfjorden. Dette gjør at det er vanskelig å skille mellom en generell nedgang over tid og samvariasjon med kovariatene, og det kan være noen få avvikende observasjoner som bestemmer om det er tid eller kovariater som fremtrer som minst viktig på de forskjellige trimmene i modelltilpasningen. Resultatene er derfor usikre indikasjoner på at det er en sammenheng med kovariater som kan utnyttes, og konklusjonen blir at det tryggeste er å basere design av prøvetakingsprogrammet på våtvektskonsentrasjonene som de er.

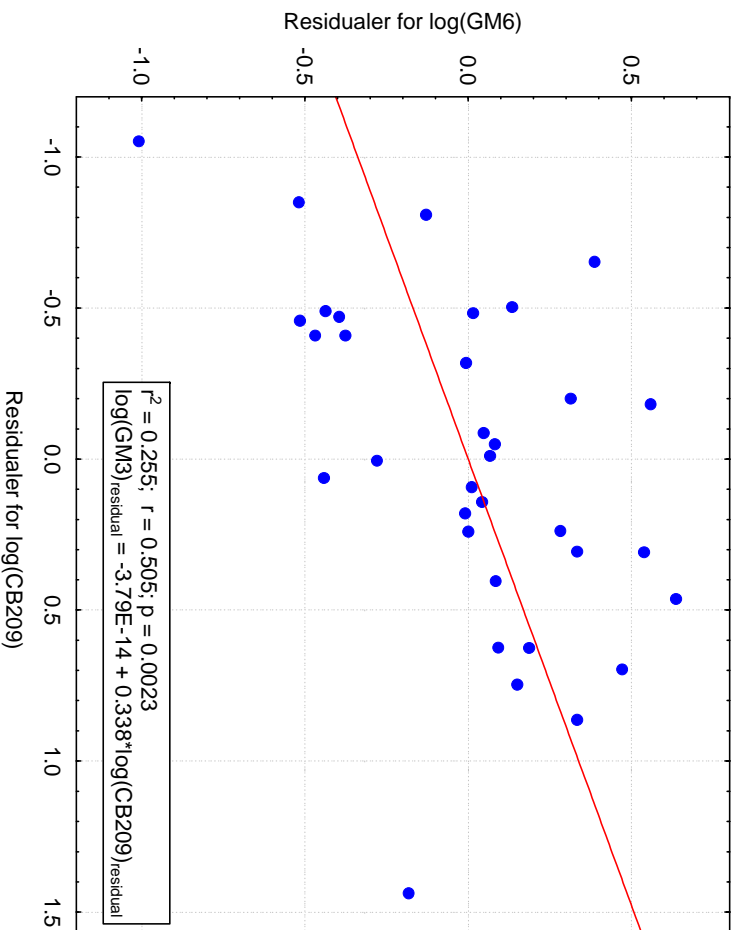
<sup>1</sup> Vekt er sterkt korrelert med lengde, og brukes derfor ikke som selvstendig kovariat. I en mer nøyaktig analyse av tidstrender og sammenheng med kovariater kunne en bruke avvik fra normal vekt ut fra lengden som egen variabel, men det anses ikke nødvendig her i en analyse av varianskomponenter som uansett er svært usikkert og hvor det bedre å være konservativ mht. muligheten til å redusere variansen.

### Andre klororganiske miljøgifter

For forskelever fra Grenland finnes det for en del prøver også data for innhold av HCB, OCS og DCB, som også er bestemt i de fleste prøvene fra Kristiansandsfjorden. Dette gjelder alle prøver fra Frierfjorden, Breviksfjorden og Såstein (Langesundsbukta) fra 1991 til og med 2001. For 2002 og 2003 mangler data for HCB, OCS og DCB.

For å se sammenhengen mellom dioksinkonsentrasjoner og kovariater er det først gjort kovarians-analyser (via Statistica GLM-modul) på konsentrasjoner av dioksin og HCB, OCS og DCB med delområde som kategorisk faktor og år som kovariat. Det er lagt inn interaksjon mellom delområde og år, slik at regresjonskoeffisienten med tid bestemmes uavhengig for hvert område. For de kombinerte dioksinmålene på våtvektsbasis gir analysen en reduksjon med ca. 10 % pr. år for Breviksfjorden og Langesundsbukta, signifikant med  $p=0.002$  til 0.03, og svakere nedgang i Frierfjorden på ca. 6 % pr. år som bare er marginalt signifikant ( $p=0.08$ ). Reduksjonstraten er ikke signifikant forskjellig mellom områdene, men interaksjonen ble likevel beholdt for å sikre at residualene ikke har noen systematisk variasjon med tid som forstyrrer studiet av sammenhengen mellom residualene for ulike variable. For HCB, OCS og DCB er bildet blandet. HCB viser signifikant nedgang med tid i alle tre områder med 14-23 % pr år, mens det svakere reduksjon for OCS og DCB.

Deretter er samvariasjon mellom residualene fra kovariansmodellen for dioksinvariable og de andre klororganiske stoffene undersøkt. Det viser seg at det stort sett ikke er noen tegn til sammenheng med HCB eller OCS, men derimot til dels ble funnet signifikante sammenhenger med DCB (CB209). Nærmere inspeksjon av resultatene viser imidlertid at sammenhengene ikke er så sterke, og at de er sterkt influert av noen få observasjoner. Et eksempel er vist i **Figur 8** hvor en positiv sammenheng fremtrer som sterkt signifikant ( $p=0.0023$ ) til tross for en sterkt avvikende verdi motsatt vei. Sammenhengen reduserer variansen med 25 % ( $r^2=0.25$ ), men figuren viser en svært ujevn fordeling rundt regresjonslinjen, slik at de matematiske forutsetningene for de statistiske parametrene neppe er oppfylt. Den signifikante regresjonen skyldes innflytelsen fra noen få prøver med lave residualer – det er ikke noen tendens til sammenheng for høye residualer. Dette er ikke noe robust resultat som kan forventes å gjelde også for fremtidige data. I lys av de store usikkerhetene i variansestimaterne kan det ikke anses hensiktsmessig å legge inn i vurderingen av utsagnskraft og behovet for prøvetaking at en skal kunne analysere andre stoffer som erstatning for dioksin. Det utelukker ikke at en ved senere analyse av innsamlende data kan prøve å utnytte andre miljøgiftkonsentrasjoner i sammenheng med dioksindata.



**Figur 8.** Sammenheng mellom residualer fra kovariansmodell for CB209 og dioksinkombinasjon GM6 i data for torskelever fra Grenlandsområdet.

## 6.2 Data fra Kristiansandsfjorden

### 6.2.1 Beskrivelse av data

Dioksiner i torsk er analysert i 19 blandprøver fra Kristiansandsfjorden, dvs. samlinger av individer fanget innenfor gitt område og tidsrom (stort sett innenfor 1-2 måneder). For alle individsettene er det analysert på blandprøver av lever. Det er området Vesterhavna/Fiskåbukta (område B/C)<sup>1</sup> som har best tidsmessig dekning, med prøver annet hvert år fra 1984 til 1992, og dessuten i 1996. En oversikt er vist i **Tabell 14**.

Torsk ble i 1984 fanget i februar-mars, i 1986 i juni og 1990 i mars; ellers ble fangstene tatt i oktober-desember. Stort sett er det ikke tatt flere prøver fra samme sted og tid. Det eneste mulige unntaket er høsten 1996, da det ble analysert to leverprøver fra området B/C: en fra fangst ved Dybingen og en fra området ved Bragdøy som er nokså nær. Begge prøver ble analysert av NILU. I 1992 ble det gjort parallell dioksinanalyse av NILU og IFHH på deler av samme leverprøve fra område D, og her brukes bare tallene fra NILU i den statistiske analysen.

<sup>1</sup> For fisk er det i tidligere rapporter bedømt å ikke være grunnlag for å skille mellom område B og C (se for eksempel Knutzen m.fl. 1991 og 1998).

**Tabell 14.** Oversikt over dioksindata for forskelever fra Kristiansandsfjorden med fordeling på stasjoner og år. Tallet for hver kombinasjon av delområde og år angir antall analyser.

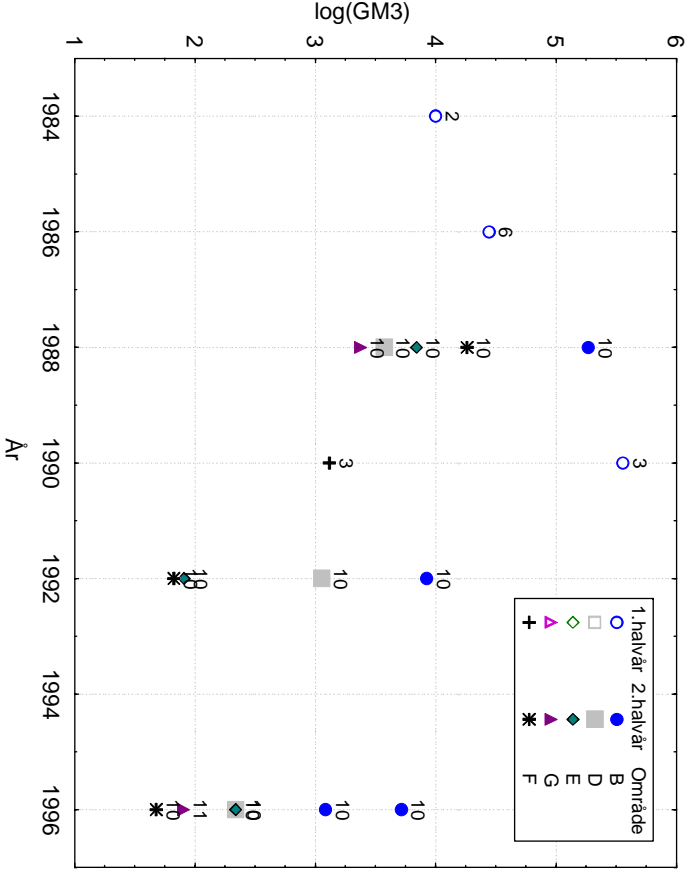
Delområde	År					
	1984	1986	1988	1990	1992	1996
	B/C	1	1	1	1	2 <sup>#</sup>
Delområde	D			1	2*	1
	E			1		1
	F			1	1	1
	G			1		1

<sup>#</sup>: analyse på to prøver

\*: Parallell-analyse av samme prøve av NILU og IFFH.

6.2.2 Statistiske analyser

Det er nokså korte tidsserier i dette materialet, og vanskelig å gjøre noen skikkelig analyse av varians-komponenter. Figur 9 viser tidsutviklingen for den kombinerte variabelene log(GM3) som ble definert i kapittel 3.6. I figuren er verdiene markert etter geografisk område og årstid, og månedsnummer er angitt for hvert punkt. Alle områdene viser generelt nedgang fra 1988 til 1996, ved at prøvene fra 1988 og 1990 har høyere verdier enn i 1992 og 1996. Det er derimot ikke så klart bilde av en videre nedgang fra 1992 til 1996. Tidsutviklingen gir visuelt inntrykk av å være ulik i de forskjellige områdene med en utflatting i 1992-1996 i område E og F og mer jevn nedgang i tidsrommet 1988 til 1996 for område B og D, men det er ikke mulig å fastslå om dette inntrykket skyldes tilfeldige utslag av variasjon mellom prøver eller faktiske trender. Når det gjelder den tidligste perioden fra 1984 til 1988 i området B/C, viser data faktisk en økning; det er motsatt av blåskjelldata fra samme område. Bildet blir omtrent det samme om vi ser på de viktigste enkeltkomponentene eller andre kombinasjonsverdier.



**Figur 9.** Tidsutvikling for dioksin i torskelerver fra Kristiansandsfjorden, geometrisk middel av tre komponenter GM3. Årstid angitt med symboltype – måned vist som tall over hver verdi.

For å få et estimat for residualvarians fra dette datasettet er det gjort en variansanalyse med år og område som kategoriske hovedeffekter, og interaksjon år\*område slått sammen med varians mellom replikater til et felles estimat for residualvariasjon. Resultatene er vist i Tabell 15. Siden det bare er ett tiffelle av replikater, vil varians for interaksjonen område\*år i hovedsak være en sum av varians for variasjon i glemmonsnitt over tid og utvalgsusikkerhet for replikate prøver ut fra variasjon mellom individer. Den residualvariansen som anslås her vil også inkludere ulikheter i tidstrend mellom områder og er sannsynligvis høyere enn det en vil få ved en analyse på et større materiale, hvor det er mulig å ha ulike modeller for tidstrend på ulike stasjoner. Den ligger også noe høyere enn estimatene for Grenlandsmaterialet. Det kan likevel indikere hva en vil måtte regne med på et begrenset datamateriale, som det som nå foreligger fra Kristiansandsfjorden.

**Tabell 15.** Residualvarians fra variansanalyse på data for dioksin i torskelerver i Kristiansandsfjorden

Analyisert variabel	Residualvarians med 9 frihetsgrader	
	Estimat	80 % konf.intervall
log(2378-TCDF)	<b>0.20</b>	0.12 0.43
log(12378/12348-PeCDF)	<b>0.30</b>	0.18 0.65
log(23478-PeCDF)	<b>0.29</b>	0.17 0.62
log(123478/123479-HxCDF)	<b>0.70</b>	0.43 1.52
log(123678-HxCDF)	<b>0.56</b>	0.35 1.22
log(234678-HxCDF)	<b>0.64</b>	0.39 1.39
log(TEE3)	<b>0.22</b>	0.14 0.48
log(Sum3)	<b>0.23</b>	0.14 0.50
log(GM3)	<b>0.20</b>	0.12 0.44

Lengde, vekt og fett-% i lever gitt for alle unntatt en prøve fra 1984. Det er forsøkt å gjøre analyser hvor ulike kombinasjoner av disse variablene er tatt med som forklaringsvariable i tillegg til år og stasjon, men det gir stort sett ikke signifikante resultater. Det utelukker ikke at det er slike sammenhenger, men betyr at datamaterialet er for spinkelt til å påvise det.

6.3 Sammenfattende konklusjon for torskelerver

Resultatene fra Grenland indikerer en varians på 0.07 for uregelmessig reelle flukturasjoner over tid for de logtransformerte kombinerte dioksimålene. I hovedsak er dette antagelig uavhengig for ulike delområder. I tillegg kommer en utvalgsusikkerhet med varians 0.09 på logskala mellom replikate blandprøver med 20 individer. Samlet varians for uregelmessig variasjon mellom enkeltprøver fra forskjellig delområde og år blir estimert til summen av dette, dvs. 0.16 på logskala. Resultatene fra Kristiansandsfjorden stemmer godt med dette. Det er ikke tegn til noe vesentlig uregelmessig flukturasjon mellom år som er felles for delområdene, dersom det er en slik effekt har den antagelig mye mindre varians enn 0.02.

Midlere reduksjonsrate er estimert til ca. 8 % pr. år for konsentrasjoner i torskelerver fra Grenland i en 10-års periode etter de store utslippsreduksjonene. Datamaterialet fra Kristiansandsfjorden er for spinkelt til å gi noe brukbart estimat for tidstrend, og et slikt estimat ville uansett bygge på data for en periode med fortsatt utslippsreduksjon. Estimatet fra Grenland antas realistisk også for en utvikling etter utslippsreduksjoner i Kristiansandsfjorden.

Det er undersøkt om biologisk karakteristikk, som lengde, vekt eller fett-% kan brukes som kovariater for å justere konsentrasjonene for naturlig biologisk variasjon. Resultatene gir bare usikre indikasjoner på at det er en slik sammenheng, og det tryggeste er derfor å basere design av prøvetakingsprogrammet på våtvektkonsentrasjonene som de er.

Det er også underøkt om empirisk sammenheng med andre klororganiske miljøgifter som er billigere å analysere kan brukes til å effektivisere overvåkingen av dioksiner. Det er ikke funnet noe robust resultat som kan forventes å gjelde også for fremtidige data, og det anses derfor ikke hensiktsmessig å legge inn i vurderingen av utsagnskraft og behovet for prøvetaking at en skal kunne analysere andre stoffer som erstatning for dioksin. Det utelukker ikke at en ved senere analyse av innsamlende data kan prøve å utnytte andre miljøgiftkonsentrasjoner i sammenheng med dioksindata.

De statistiske analysene av tidstrender for dioksininnhold i torskelever basert på datamaterialet fra Kristiansandsfjorden og Grenlandsfjordene tilsier:

- Hvis tidsutviklingen i dioksinkonsentrasjoner på ett sted overvåkes ved at en tar prøver av torskelever over en årrekke, med 20 fisk i hver prøve, vil konsentrasjoner for enkeltprøver variere tilfeldig fra år til år rundt jevn tidstrend. Variasjonene har relativt standardavvik ca. 40 til 50 % (varians 0.16 for naturlige logaritmer av konsentrasjon).
- Noe over halvparten av variansen ser ut til å være utvalgsusikkerhet pga. tilfeldig variasjon mellom individuelle torsk fra samme sted og tid (varians 0.09 for naturlige logaritmer av konsentrasjon). Resten skyldes antagelig reelle avvik fra trenden i gjennomsnittet fra år til år (varians 0.07 for naturlige logaritmer av konsentrasjon). Hvis det analyseres flere prøver fra samme stasjon hvert år, eller hvis antall individer pr. prøve økes, vil virkningen av utvalgsusikkerheten avta, mens virkningen av de reelle avvikene vil være den samme.
- Hvis det tas parallelle tidsserier fra flere stasjoner innenfor et område, kan en regne med at disse avvikene fra tidstrendene er uavhengige på hver stasjon. Usikkerheten i samlet års-gjennomsnitt vil da minske jo flere stasjoner en har. I tillegg kan det være parallelle avvik som er felles for alle stasjoner innenfor et område (estimert varians 0.02 på naturlig logskala ut fra Grenlandsmaterialet). Et slikt variansbidrag vil ikke ha særlig betydning totalt sett hvis antall stasjoner er 4 eller mindre.
- Data fra Grenland viser ca. 8 % reduksjon av konsentrasjoner i torskelever pr. år over en 10-års periode etter store utslippsreduksjoner. Det antas at det er realistisk å vente det samme i Kristiansandsfjorden etter de store utslippsreduksjonene som har skjedd der siden 1990.
- Det er ikke funnet sikre sammenhenger med biologisk karakteristikk eller andre klororganiske miljøgifter, og analyse av utsagnskraft bør ikke baseres på at det er mulig å utnytte dette.

# 7 Statistisk analyse av data for filét av skrubbeflyndre

For skrubbeflyndre er det stort sett bare filét som er analysert; data for lever finnes bare for noen få prøver. Bare data for filét analyseres her.

## 7.1 Grenland

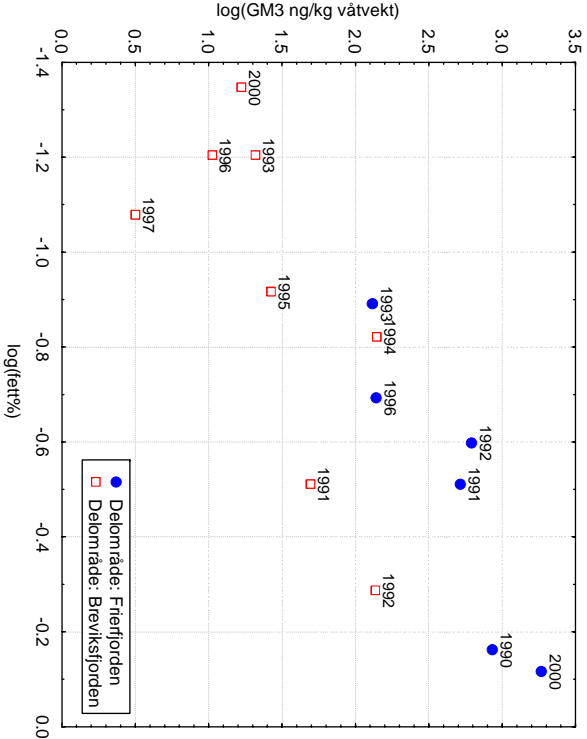
### 7.1.1 Beskrivelse av data

Fra Grenland er det analysert prøver av skrubbeflyndre fra Frierfjorden og Breviksfjorden med tidsmessig deking som vist i **Tabell 16** nedenfor. Prøvene fra Frierfjorden i 1987 og 1990 er tatt om høsten (hvv. august og oktober); alle de andre prøvene fra april eller mai. Det er ingen replikate prøver fra samme sted og år, men en viss overlapping i tid mellom de to tidsseriene. De fleste prøvene er blandprøver av ca. 20 individer, men noen bare av ca. 10 individer.

**Tabell 16.** Antall prøver av skrubbefilét fra Grenland fordelt på område og år.

Delområde	1987	1990	1991	1992	1993	1994	1995	1996	1997	2000
Frierfjorden	1	1	1	1	1			1		1
Breviksfjorden			1	1	1	1	1	1	1	1

Våtvektskonsentrasjonene viser en generell nedadgående trend i begge områder, med sterkt avvikende høye verdier i prøven fra Frierfjorden i år 2000. Ved en tidligere analyse av den vanlig brukte toksisitetsekvivalenten for dioksin på dette materialet ble det imidlertid funnet en ganske klar proporsjonalitet med fett-% innenfor hvert område (Bjerkeng og Ruus 2002). Det samme gjelder den kombinerte dioksinvariabelen log(GM3) (se **Figur 10**).



**Figur 10.** Dioksinkonsentrasjoner i skrubbefilét fra Grenland som beregnet kombinasjonsmål GM3 som funksjon av fett-%, med markering av år og område. Begge akser er på logaritmisk skala.

Prøvene fra begge områdene har tendens til minskende fett-% med tid gjennom perioden 1990-2000, med unntak av en høy fett-% nettopp i prøven fra Frierfjorden i 2000 som stemmer godt med den høye våtveikonsentrasjonen for dioksin. Når konsentrasjonene justeres for fettinnhold, blir forskjellen mellom områdene mindre i størrelse, men likevel vel så tydelig, mens det ikke gjenstår noen signifikant reduksjon over tid i løpet av perioden 1990 til 2000. Prøven fra Frierfjorden i 1987 har den høyeste konsentrasjon på våtvekt-basis. Her mangler fettvektbestemmelse, men prøven har den største konsentrasjonen også på fett-basis, selv hvis en bruker den høyeste fett-% som ellers er målt i skrubbeprøvene. Det er valgt å se bort fra data for 1987 og 1990 fordi de er fra en annen årstid enn de andre og fordi det er best å avgrense data til de som stammer fra perioden etter siste store utslippsreduksjon for å estimere aktuelle variansmønstre ved fremtidig overvåking.

## 7.1.2 Statistiske analyser

Følgende statistiske analyser er gjort på skrubbedata fra Grenland:

- A. En kovariansmodell med område og år som tilfældige, kategoriske faktorer og  $\log(\text{fett-\%})$  som kovariat. Interaksjonen område\*år er residualledd. Det er signifikant forskjell mellom områdene, og signifikant sammenheng med fett-%. Når data begrenses til de 5 årene som har data fra begge områder, er det ikke noe signifikant variansbidrag mellom år som er felles for de to områdene. Når alle data etter 1987 tas med, blir det en signifikant variasjon mellom år for noen komponenter (0.1 for  $\log(\text{GM3})$ , men det skyldes nok at variasjon mellom år da er blandet sammen med uavhengig variasjon innen hvert område. Residualleddet estimeres bare ut fra de 5 årene hvor det ble tatt prøve fra begge områder.

- B. En alternativ modell hvor år brukes som kovariat i en log-lineær regresjon i stedet for som kategorisk faktor. Når alle data etter 1987 er med i analysen, blir residualvariansen nå bestemt med 9 frihetsgrader og vil også inkludere avvik fra år til år i forhold til en log-lineær tidstrend, for enkelstasjon i de årene hvor det bare er tatt prøve i ett område, og som gjennomsnitt over stasjonene for de årene som har en prøve fra hvert område. Residualvariansen blir da også større enn for analyse A, men det kan også skyldes at flere data er med i grunnlaget, og betyr ikke at det er noen signifikant felles variasjon mellom år som gjennomsnitt over de to områdene. Det er signifikant sammenheng med fett-% og område for de fleste komponentene. Log-lineær regresjon mot år er ikke signifikant i forhold til de uregelmessige svingningene men har et tosidig 80 % konfidensintervall fra +0.05 til -0.06 på naturlig logskala. Resultatet utelukker altså ikke en økning over tid, men sett i sammenheng med de andre resultatene er dette lite trolig. Hvis vi tar det for gitt at det har skjedd en reduksjon, kan vi med 90 % konfidens sette en øvre grense på 6 % pr. år, altså sannsynligvis svakere enn for torskelever.

Begge modellene gir omtrent samme relasjon med fett-%. Det er signifikant korrelasjon med  $\log(\text{fett-\%})$  for de fleste komponenter, bla. for  $\log(\text{GM3})$  hvor regresjonskoeffisientene for  $\log(\text{fett-\%})$  som kovariat er hhv.  $0.85 \pm 0.21$  og  $1.18 \pm 0.35$  fra de to analysene. Også når vi ser bort fra Frierfjordprøven fra 2000 består disse trekkene, men med noe svakere relasjon til fett-%.

**Tabell 17** viser estimerte varianskomponenter fra de forskjellige analysene. Residualvariansen fra analyse A er konsekvent mindre enn fra analyse B, men de fleste konfidensintervallene er overlappende. Forskjellen kan derfor skyldes ulikt datagrunnlag, og ikke at det er noen reell variasjon fra år til år som er felles for de to områdene. Alle variansestimaterne er ganske usikre, så konklusjonene er beheftet med stor usikkerhet, men det beste estimatet gir en samlet varians mellom prøver tatt i forskjellige år på 0.14 for kombinasjonsmålene. Det er ikke mulig å antyde noen oppdeling i bidrag fra varians for ending av gjennomsnitt over tid og utvalgusikkerhet.



**Tabell 17.** Varianskomponenter for skrubbefillet fra Frierfjorden og Breviksfjorden ut fra variansanalysene.

Analysert variabel	Residualvarians fra analyse A bestemt med 3 frihetsgrader		Residualvarians fra analyse B bestemt med 9 frihetsgrader	
	Estimat	80 % konf.intervall	Estimat	80 % konf.intervall
log(2378-TCDF)	0.148	0.07 0.76	0.21	0.13 0.46
log(12378/12348-PeCDF)	0.020	0.01 0.10	0.13	0.08 0.28
log(23478-PeCDF)	0.008	0.00 0.04	0.14	0.09 0.31
log(123478/123479-HxCDF)	0.025	0.01 0.13	0.11	0.07 0.23
log(123678-HxCDF)	0.023	0.01 0.12	0.13	0.08 0.27
log(234678-HxCDF)	0.173	0.08 0.89	0.30	0.18 0.65
log(TE3)	0.013	0.01 0.06	0.14	0.09 0.30
log(SUM3)	0.046	0.02 0.24	0.15	0.09 0.33
log(GM3)	0.031	0.01 0.16	0.14	0.09 0.31

7.1.3 Sammenheng med andre miljøgifter

Det undersøkt om residualene fra kovariansmodellen for log(GM3) kan korreleres med tilsvarende residualer for log(HCB), log(OCS) og log(DCB). Den beste sammenhengen finnes med DCB. Residualene for log(GM3) og log(DCB) er korrelert med regresjonskoeffisient 0.15, og variansen etter justering for DCB reduseres med ca. 35 %. Hele sammenhengen er imidlertid knyttet til bare to av 10 prøver, dvs. at det er lite grunn til å tro at sammenhengen skal gjelde for fremtidige prøver.

7.2 Kristiansandsfjorden

7.2.1 Beskrivelse av data

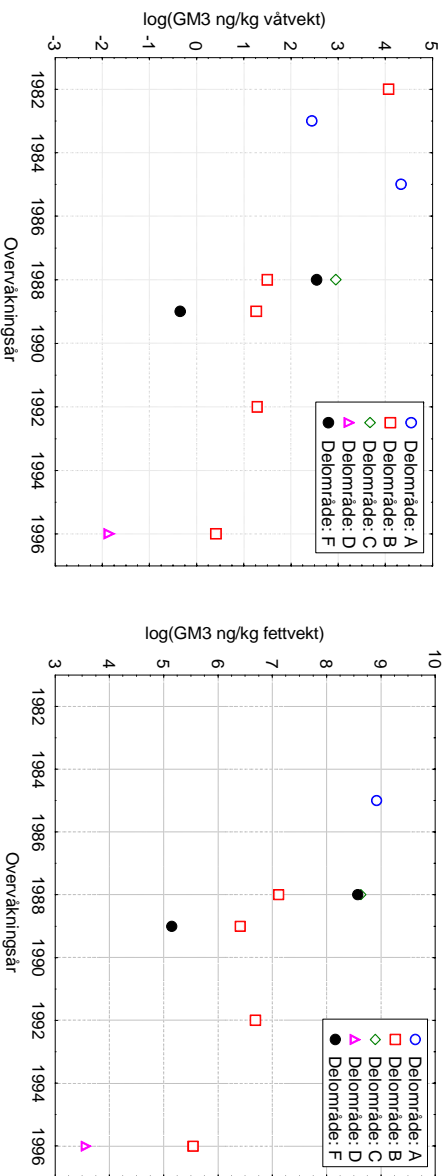
Dioksin er analysert på i alt 13 prøver av skrubbeflyndre. Stort sett En oversikt over data er gitt i

**Tabell 18.** Fangstene er gjort til ulik tid på året: Juli 1982, februar 1984, juni 1986, oktober 1988 og 1996, og mars 1990. Prøvene er blandprøver med opp mot 15-20 fisk – det er oppgitt at noen prøver er basert på svært få individer.

**Tabell 18.** Oversikt over dioksindata for fillet av skrubbeflyndre fra Kristiansandsfjorden fordelt på stasjoner og år. Det finnes bare en prøve for hver kombinasjon av vev, sted og år.

Delområde	Stasjonsnavn	1982	1984	1986	1988	1990	1992	1996
A	Hannevigsbukta		1	1				
B		1			1	1	1	1
C	Fiskåbukta				1			
D	Fossevika							1
F	Randøy/Kalvøy				1	1		

Her er det svært få prøver, og det er bare fra område B at det er noen form for tidsserie. Plott av tidsforløpet for de forskjellige komponentene viser en generell fallende trend, med noe variasjon mellom komponenter. Som eksempler er vist log(GM3) både på vårvекtsbasis og fett-basis (**Figur 11**).



**Figur 11.** Dioksinkonsentrasjon log(GM3) på våtvekt- og fettvekt-basis i skrubberfjndre fra Kristiansandsfjorden

7.2.2 Statistiske analyser

Generelt er det ikke grunnlag for å si noe om forskjell mellom områdene, alle data gir et samlet bilde av en fallende tidstrend, men med relativt stor spredning, og det skal ikke legges for stor vekt på den jevne nedgangen over tid i område B for noen dioksivariable og for log(G3) som vist i figurene. Materialet er også for spinkelt til å skille ut noen signifikant sammenheng med fett-%, eller se systematisk på forskjell mellom områder. I stedet er det gjort to enkle regresjonsanalyser mot overvåkingsår for å antyde mulig variasjonsområde for residualvarians (se **Tabell 19**). Analysene er gjort på log(GM3) normalisert til fett-innhold, noe som stemmer med sammenhengene som ble funnet for Grenlandsdata.

**Tabell 19.** Resultat av lineær regresjon mot overvåkingsår for log(GM3) på fettvektsbasis i skrubberfjlet fra Kristiansandsfjorden, alle data fra og med overvåkingsår 1985

	Regresjonskoeffisient mot overvåkingsår med st.avvik	residualvarians			
		estimat	frihets- grader	80 % konf.intervall	
Uten å skille mellom områdene	-0.38±0.11	1.36	7	0.8	3.4
Bare data for område B	-0.16±0.06	0.16	2	0.07	1.5

Analysen av alle områdene sett under ett influeres sterkt av forskjellen mellom data fra ulike områder, og er neppe representativ for overvåking med samme geografiske dekning over tid. Analysen på data bare fra område B gir også usikre estimater, men her stemmer regresjonskoeffisienten med det som er funnet for andre prøvetyper, og residualvariansen har et konfidensintervall som overlapper godt med de intervallene som ble estimert for Grenland.

### 7.3 Sammenfattende konklusjon for skrubbeflyndre

Datamaterialet for skrubbeflyndre er relativt spinkelt både for Grenland og Kristiansandsfjorden, og gir stort sett bare grunnlag for å estimere varians for uregelmessige fluktasjoner i forhold til tidstrender eller gjennomsnittsnivå når det tas en prøve pr. stasjon og år. Det beste estimatet for denne residualvariansen er omkring 0.15 på logtransformerte konsentrasjoner, og det ser ut til å gjelde både for Grenland og Kristiansandsfjorden. Det er ikke grunnlag for å estimere varians for parallelle avvik over flere stasjoner eller splitte opp residualvariansen i utvalgusikkerhet for enkeltprøver og reell variasjon mellom år og stasjoner, slik som for blåskjell og torsk.

De statistiske analysene av tidstrender for dioksininnhold i skrubbeflyndre basert på datamaterialet fra Kristiansandsfjorden og Grenlandsfjordene tilsier:

- Hvis tidsutviklingen i dioksinkonsentrasjoner på ett sted overvåkes ved at en tar prøver av fillet fra skrubbeflyndre over en årrekke, med 20 fisk i hver prøve, vil konsentrasjoner for enkeltprøver variere tilfeldig fra år til år rundt en jevn tidstrend. Variasjonene tilsvarer et relativt standardavvik på ca. 40-50 % for enkeltkonsentrasjoner (variens 0.15 for naturlige logaritmer av konsentrasjoner).
- Avvikene skyldes delvis utvalgusikkerhet pga. tilfeldig variasjon mellom enkeltfisk fra samme sted og tid, men noe kan også skyldes reelle avvik i gjennomsnittet fra år til år i forhold til langsiktig trend. Datagrunnlaget for skrubbefillet er ikke stort nok til at en kan tallfeste betydningen av utvalgusikkerhet i forhold til reelle avvik.
- Hvis det tas prøver fra flere delområder hvert år, kan avvikene antas å være uavhengige for hvert delområde. Usikkerheten for samlet årsgjennomsnitt vil da minske jo flere delområder en har prøver fra. Det er ikke mulig å påvise avvik mellom år som er felles over flere stasjoner for det eksisterende datamaterialet.
- For gitt tid og sted ser konsentrasjonene ut til å variere systematisk omtrent proporsjonalt med fett-innholdet. Standardavviket for tilfeldige avvik som er angitt ovenfor gjelder etter at det er korrigert for variasjon i fett-innholdet mellom prøvene.
- I Grenlandsområdet kan det ha vært en reduksjonsrate på opp til ca. 6 % pr. år, men mest sannsynlig mindre enn dette, og usikkerheten er for stor til at en kan påstå at konsentrasjonene i skrubbeflyndre er redusert i det hele tatt i perioden fra 1990 til 2000.

# 8 Statistisk analyse av data for krabbe

For krabbe er de fleste dioksinanalysene gjort på prøver av krabbesmør. Det finnes også noen data for kløkjøtt og rest skallinnmat, men det er bare for krabbesmør det finnes tidsserier som er brukbare for estimering av varianskomponenter. Mest data finnes fra Grenland; fra Kristiansandsfjorden er det så få prøver at det er begrenset hva en kan få ut av dem isolert sett.

De fleste analysene er gjort av NIL U. IFFH har analysert bare et par prøver fra Grenland, den ene som parallellanalyse til NIL U. Tre av prøvene fra Kristiansandsfjorden er analysert av IFFH, de andre av NIL U. Den statistiske analysen baseres bare på NIL Us data for Grenland – for Kristiansandsfjorden brukes også analysene fra IFFH, bortsett fra der det er analysert parallelt av IFFH og NIL U på samme prøvemateriale.

I kapittel 8.1 presenteres en statistisk analyse av data for Grenland. For dette datamaterialet er det aktuelt med en mer kompleks analyse enn for de andre prøvetypene, med bruk av kjønn som kovariat og estimering av flere typer variansbidrag. Det er derfor stort sett valgt å avgrense analysen til kombinasjonsverdien GM3 definert i kapittel 3.6 (geometrisk middel av tre kongenerer). De fleste resultater i dette kapitlet gjelder log(GM3). Datamaterialet fra Kristiansandsfjorden er mye mindre, og analyseres på en mye enklere måte. Det gjøres i kapittel 8.2 i lys av resultatene fra Grenland, også begrenset til log(GM3).

## 8.1 Data fra Grenland

### 8.1.1 Beskrivelse av data

Data for krabbesmør finnes i lengre tidsserier fra 6 lokaliteter i Grenlandsområdet. En oversikt er vist i **Tabell 20**. I 1998 og 1988 ble det tatt prøver fra flere lokaliteter i Jomfrulandområdet (Askholmen, Kreppa) og i Langesundsbukta (Steingrunnen, Nevlunghamn. (se **Figur 12**). Prøvematerialet i tabellen er det som er mest egnet for en statistisk analyse hvor tidstrender inngår.

**Tabell 20.** Oversikt over prøvematerialet for dioksiner i krabbesmør i Grenlandsområdet

Delområde	Stasjonsnavn	1987	'88	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	2000	'01	'02	'03
Frierfjorden	Ringsholmen	1	1	1	1	1	1	1	2*	1	1		1	1			
Breviksfjorden	Bjørkøy			1	1	1	2 <sup>#</sup>	1	2*	1	1		1	1	1	1*	1*
Langesundsbukta	Arøya	1	3*	2*	4	1	1	1	2*	1	1		1	1	1		
Langesundsbukta	Såstein	1	2*	1	4	1	1	1	2*	1	1		1	1			1 <sup>&amp;</sup>
Langesundsbukta	Åbyfjorden		1	1	1	1		1	2*	1	1		1	1	1	1	
Jomfrulandområdet	Jomfruland		1	1	1	1	1		2*	1	1			1		1 <sup>&amp;</sup>	

<sup>#</sup>: En prøve er analysert parallelt av IFFH og NIL U – bare NIL U–resultatet er tatt med i analysen.

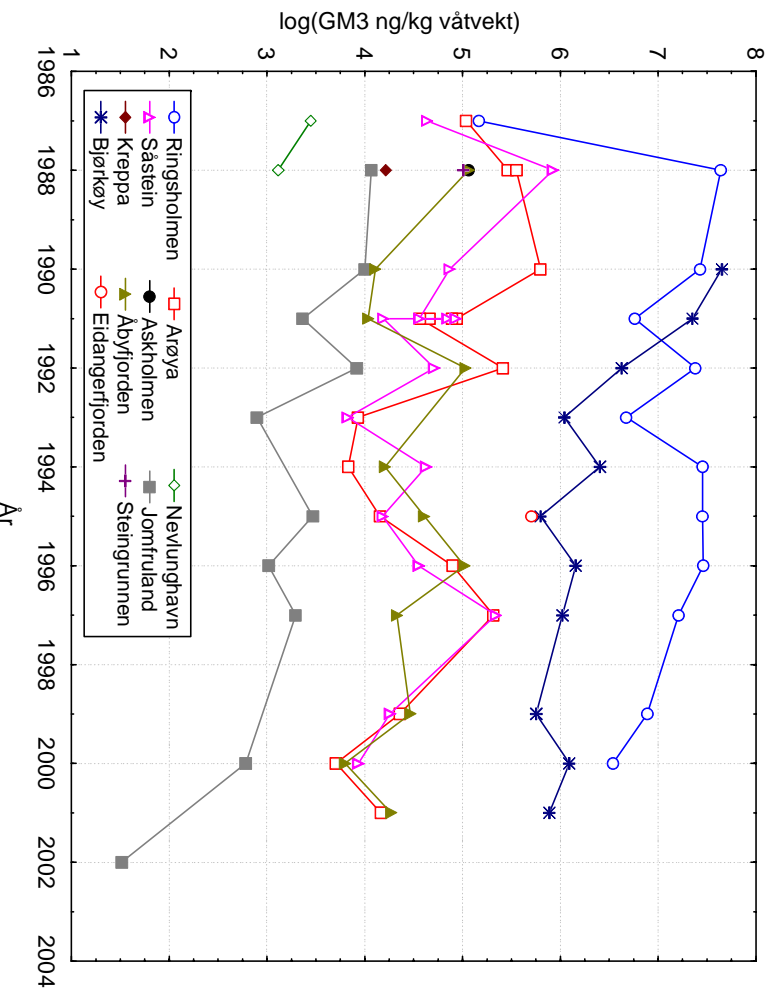
<sup>\*</sup>: En prøve av humnkrabbe

<sup>&</sup>: Blandet hann og hunn

Alle prøvene er blandprøver, for det meste av 20 individer. Med unntak av ett tilfelle består hver prøve av individene av bare ett kjønn, og de fleste prøvene er av hannkrabbe. Forekomst av prøver av humnkrabbe er avmerket i tabellen. Det finnes noen tilfeller av replikate prøver fra samme lokalitet og år, stort sett er det da en prøve av hvert kjønn, men det er også noen tilfeller av replikate prøver av hannkrabbe. Alle prøvene er tatt om høsten (august til november), slik at kalenderår og overvåkningsår

alltid er identisk. For alle krabbeprøvene er også fett-% oppgitt, slik at det kan inngå som kovariat i analysen.

**Figur 12** viser målte verdier av dioksin i krabbesmør på våvektsbasis som  $\log(\text{GM3})$ . For å unngå forstyrrelser fra forskjell etter kjønn er bare prøvene for hannkrabbe tatt med. Figuren gir visuelt inntrykk av en langsomt nedadgående tidstrend fra 1990 til 2001, omtrent parallelt på alle stasjoner med data for hele perioden. De tilfældige avvikene rundt tidstrenden er til dels også parallelle over flere stasjoner (1990-93 for de fleste stasjoner; 1999-2001 for stasjonene Arøya, Såstein og Åbyfjorden i Langesundsbukta). Prøven fra Ringsholmen i 1987 har svært lave verdier, faktisk 5 ganger lavere enn noen senere prøve fra samme sted, og omtrent på linje med konsentrasjoner fra Arøya og Såstein. Det kan se ut som det har skjedd en feil, og i alle fall er 1987-prøven fra Ringsholmen så avvikende at det er rimelig å se bort fra den i en analyse av varianskomponenter for fremtidig overvåking. Siden noe av det samme også gjelder to andre prøver fra 1987, er det valgt å se bort fra alle prøvene fra 1987 i den statistiske analysen her.



**Figur 12.** Dioksinkonsentrasjoner av krabbesmør fra Grenlandsområdet, vist med kombinasjonsmål GM3 på logskala. (enhet ng/kg våtvekt). Bare data for hannkrabbe er inkludert, og data fra 2003 er derfor ikke vist her.

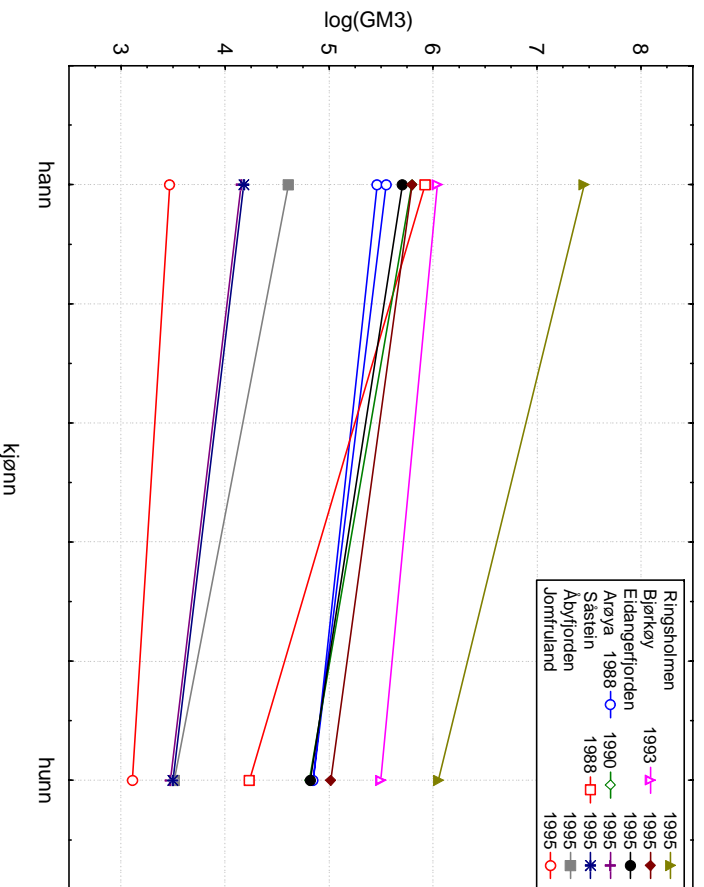
## 8.1.2 Statistiske analyser

### Forskjell mellom humn- og hann-krabbe

Som det fremgår av Tabell 20 er det ved 10 tilfeller tatt prøver parvis av både humn og humn-krabbe fra samme tid og sted. Det gir grunnlag for å undersøke om det er forskjell mellom kjønnene. Det er gjort en kovariansanalyse på disse prøvene hvor kategoriske hovedfaktorer er *kjønn* og *prøvetakingsstillelle* (=kombinasjon år og stasjon). Log(fett-%) er tatt med som kovariat, siden det i en tidligere statistisk

analyse av fullstendig toksisitetsekvivalent for det samme materialet ble funnet tegn til sammenheng med fett-%. Resultatene tydet på at  $\log(\text{TE})$  var korrelert med  $\log(\text{fett-\%})$  med regresjonskoeffisient anslått til 0.8-0.9, om enn med stor usikkerhet (Bjerkeng og Ruus 2002).

Kovariansanalysen av hann-hunn prøvepar viser en klar effekt av kjønn for alle de 6 utvalgte enkeltkomponentene og for kombinasjonsverdiene som bygger på tre komponenter. Hannkrabbe har høyere konsentrasjoner enn hunnene, og forskjellen på logskala er 0.8 til 0.9 (standardfeil  $\pm 0.11$ ) dvs. at hannene i snitt har ca. 2.5 ganger høyere konsentrasjon enn hunnkrabbe på samme sted og tid. Det gjelder altså generelt både for enkeltkomponenter og kombinasjoner av disse. **Figur 13** viser  $\log(\text{GM3})$  for de enkelte prøveparene som inngår i analysen. Analysen viser ingen signifikant sammenheng med fett-%, men det utelukker ikke at det er en slik sammenheng; 90 % konfidensintervall for regresjonskoeffisient med  $\log(\text{fett-\%})$  går fra -0.89 til +0.96, og det overlapper så vidt estimerte koeffisienter funnet av Bjerkeng og Ruus (2002). For å se om virkning av fett-% eventuelt er innebygd i forskjellen mellom kjønn er det gjort en supplerende analyse med  $\log(\text{fett-\%})$  som avhengig variabel og kjønn og prøvetakingstilfelle som faktorer. Den viser ingen systematisk forskjell i fett-% mellom hannen og hunner, så det ser ikke ut til å være slik at de to effektene er koblet i disse dataene. Residualvariansen i denne analysen er ca. 0.07 for  $\log(\text{GM3})$ , bestemt med 11 frihetsgrader. Den er et estimat for hvor mye forskjellen mellom prøver av hann- og hunnkrabbe fra samme tid og sted varierer, og kan tolkes som en utvalgsusikkerhet for enkeltprøve.



**Figur 13.** Forskjell mellom dioksininnhold i hann- og hunnkrabbe i Grenland, ved parallelle prøvetakinger på samme tid og sted, som  $\log(\text{GM3})$  på våtvektsbasis.

## Variansanalyse på utvalgte deler av datasettet for hannkrabbe

For å se på forskjeller mellom år og stasjoner, og interaksjonen mellom dem, kan en se på to delvis overlappende utvalg av data med fullstendig dekning av alle kombinasjoner av år og stasjoner, hvor bare data for hannkrabbe er med:

- A. Data fra år 1990-1997, 1999 og 2000 for stasjonene Ringsholmen, Bjørkøy, Arøya og Såstein.
- B. Data bare fra årene 1990, 1992, 1995-1997 og 2000, til gjengjeld er også stasjonene Åbyfjorden og Jomfruland tatt med i tillegg til de som er nevnt under A.

Kovariansanalyse er gjort på disse to datasettene med stasjon som fast faktor, år som tilfeldig faktor og inkludert interaksjonsledd år\*stasjon. Resultatet blir omtrent det samme for begge datasettene. For  $\log(\text{GM3})$  blir variansestimater for fluktasjoner mellom år, felles over stasjoner, hhv. 0.145 og 0.11 på logskala. Dette omfatter både langsiktig trend og uregelmessig variasjon fra år til år rundt trenden. Variansestimater for interaksjonsleddet År\*stasjon er hhv. 0.092 og 0.11 for de to datasettene. Residualvariansen for variasjon mellom replikate prøver er estimert til 0.072 med 6 frihetsgrader, bygd på replikatene i 1991 for Arøya og Såstein. Det er forsøkt med  $\log(\text{fett-}\%)$  som kovariat også for disse datasettene, både med og uten interaksjonsledd år\*stasjon inkludert i modellen, men det er ikke mulig å finne veldefinerte sammenhenger.

### Kovariansanalyse på hele materialet

Ut fra **Figur 12** er det ikke sterke tegn til ulik tidsutvikling på forskjellige stasjoner, og heller ikke for avvik fra log-lineær tidstrend, selv om det er tegn til parallelle variasjoner for Arøya og Såstein i Langesundsbukta. En kovariansanalyse av  $\log(\text{GM3})$  er gjort på alle data unntatt de fra 1987 hvor fett-% mangler. Stasjon er kategorisk faktor og år, kjønn<sup>1</sup> og  $\log(\text{fett-}\%)$  inngår som kovariater. Med samme regresjon mot tid for alle stasjoner kan modellen skrives:

$$\log(\text{GM3}) = \mu + \alpha_{\text{stasjon}} + b_{\text{år}} \cdot \text{År} + b_{\text{jeit}} \cdot \log(\text{fett}\%) + c \cdot (\text{andel hanner}) + D_{\text{prøve}(\text{stasjon}, \text{år})} \quad (3)$$

I denne modellen blir alle regresjonskoeffisientene signifikante. Regresjonskoeffisientene blir:

$$\begin{aligned} b_{\text{år}} &= -0.088 \pm 0.012 \\ b_{\text{jeit}} &= 0.41 \pm 0.16 \\ c &= 0.79 \pm 0.14 \end{aligned}$$

Sammenheng med fett-% ligger innenfor det usikkerhetsintervallet som ble anslått ovenfor ved analyse bare på prøvepar av hann og hunnkrabbe. Anslått sammenheng med fett-% vil kunne influeres av eventuell tidsutvikling eller stasjonsforskjell i fett-% og behøver ikke nødvendigvis uttrykke en funksjonell sammenheng som også gjelder for replikate prøver fra samme tid og sted. En inspeksjon av tidsplott av fett-% og konsentrasjon av dioksin over tid pr. stasjon tyder ikke på at det er noen særlig tidstrend i fett-% sammenlignet med svingninger frem og tilbake, eller at det er særlig samsvar mellom forskjell i konsentrasjon og fett-% mellom stasjoner, så det kan være grunn til å tro at koeffisienten  $b_{\text{jeit}}$  gir et noenlunde riktig bilde av en funksjonell korrelasjon.

Det primære i denne sammenhengen er uansett residualene, dvs. avvik mellom observert verdi og modellverdi for  $\log(\text{konsentrasjon})$  korrigert for kjønn, fett-%, stasjonsforskjeller i gjennomsnitt over tid og langsiktig log-lineær tidstrend. Residualene  $D_{\text{prøve}(\text{stasjon}, \text{år})}$  fra kovariansanalysen omfatter

---

<sup>1</sup> Kjønn er kodet hann=1, hunn=0 og blandet=0.5

uregelmessige svingninger fra år til år, både felles for stasjoner og uavhengig fra stasjon til stasjon, og i tillegg varians mellom replikater fra samme år og stasjon. Residualene er analysert i en variansmodell hvor år og stasjon inngår som kategoriske faktorer, og hvor interaksjonsleddet År\*stasjon er med som spesifisert effekt i modellen, slik at det blir skilt mellom interaksjon og varians mellom replikater tatt på samme sted og tid. Modellen for denne analysen er

$$D_{pr\ddot{o}ve(stasjon,\ddot{a}r)} = \tau_{\ddot{a}r} + \delta_{stasjon*\ddot{a}r} + \varepsilon_{pr\ddot{o}ve(stasjon,\ddot{a}r)} \tag{4}$$

ANOVA-resultater og estimerte varianskomponenter for de forskjellige leddene er vist i **Tabell 21**. Både hovedeffekt år og interaksjonen stasjon\*år er signifikante; hovedeffekt år er signifikant både i forhold til interaksjonsleddet og residualvariansen. Modellen gir som rimelig er ikke signifikant forskjell mellom middelvei for stasjoner, kjønn eller fett-%, siden residualene allerede skal være korrigert for dette<sup>1</sup>, men de er likevel tatt med for å gi korrekt antall frihetsgrader. De er ikke vist i tabellen, og resultatet endrer seg ikke mye om en tar ut de ikke-signifikante effektene.

**Tabell 21.** Resultat av variansanalyse på log(GM3) justert for kjønn, stasjon og log-lineær tidstrend  
df= antall frihetsgrader (degrees of freedom), MS = midlere kvadratavvik (Mean square)

Effekter		Signifikanstest						Varianskomponent (variansbidrag)			
		Effekt		Feilledd		F				p	
		df	MS	df	MS						
Hovedeffekt År	test mot residual			17	0.081	4.49	0.0023	0.042	$\sigma^2_{år}$		
	test mot interaksjon	13	0.366	55.9	0.157	2.32	0.015				
Interaksjon År*Stasjon		49	0.170	17	0.081	2.08	0.049	0.074	$\sigma^2_{år, stasjon}$		
Residual (replikat-prøver)		17	0.081					0.081	$\sigma^2_{repl}$		

Forventningsverdien for midlere kvadratavvik (MS) for de forskjellige effektene er kombinasjoner av ulike varianskomponenter som beskrevet i Vedlegg F. Estimert MS for interaksjon år\*stasjon inkluderer restvirkingen av varians mellom enkelprøver (replikater) i tillegg til den reelle variasjonen mellom ulike kombinasjoner av år og stasjon, og estimert MS mellom år inkluderer restvirking av begge de to andre leddene. **Tabell 22** viser uttrykkene for MS som resultat av variansanalysen på residualene fra kovariansanalysen for datasettet for krabbesmør. Datasettet er ubalansert og mangle kombinasjoner av stasjon og år mangler data, så koeffisientene avviker fra det en vil vente ut fra antall frihetsgrader.<sup>2</sup>

<sup>1</sup> Hvis datasettet hadde vært balansert og med full dekning av samme år for alle stasjoner, skulle midlere forskjell mellom stasjoner og kjønn være eksakt null.

<sup>2</sup> Verdier fra Statistica 7.1, GLM-modulen i tabell for 'Expected Mean Square Coefficients'



**Tabell 22.** Uttrykk for forventet midlere kvadrat (*Mean Square*: MS) for effekter år, år\*stasjon og replikater innenfor år og stasjon i variansanalyse på hele materialet for krabbe med korreksjon for kjønn og fett-%.

	Uttrykk for <i>mean square</i> (MS) i ANOVA-tabell	frihetsgrader
Fluktuasjon i årsmidler	$EMS_{\bar{A}r} = 4.94\sigma_r^2 + 1.02\sigma_{st}^2 + \sigma_r^2$	13
Uavhengige fluktuasjoner for år*stasjon (interaksjon)	$EMS_{\bar{A}r, stasjon} = 1.185\sigma_{st}^2 + \sigma_r^2$	49
Variasjon mellom replikater	$EMS_r = \sigma_r^2$	17

Ved å kombinere uttrykkene for de ulike MS-estimaten som en vektet sum av varianskomponenter kan en da beregne varianskomponentene. Det er gjort i **Tabell 21**, hvor også tilnærmet antall frihetsgrader og 80 % konfidensintervall for varianskomponentene er beregnet. Hvordan det er gjort er beskrevet nærmere i Vedlegg F.

**Tabell 23.** Beregnede varianskomponenter ut fra variansanalyse på log(GM3) justert for kjønn, stasjon og log-lineær tidstrend.

Effekter	Varianskomponent (variansbidrag)	effektivt antall frihetsgrader	80 % konfidensintervall	
			10 % fraktil	90% fraktil
Hovedeffekt År	$\sigma_r^2$	4	0.011	0.08
Interaksjon År*Stasjon	$\sigma_{st}^2$	7	0.03	0.12
Residual (replikat-prøver)	$\sigma_r^2$	17	0.05	0.12

## 8.2 Data fra Kristiansandsfjorden

### 8.2.1 Beskrivelse av data

Som **Tabell 24** viser, er det er ganske få data for krabbesmør fra Kristiansandsfjorden, 8 prøver av hannkrabbe, 2 av hunnkrabbe og tre blandet hanner og hunner. I 1992 ble det tatt parallelle prøver av hunnkrabbe og hannkrabbe fra to områder, og dette lille datasettet kan undersøkes for å se om forskjell mellom kjønn stemmer med det som ble funnet for Grenland. I 1988 og 1989 var 3 av 7 prøver en uspesifisert blanding av hanner og hunner. Holder vi blandingsprøvene utenfor, har vi et balansert datasett bestående av prøver av hannkrabbe fra område B og E i 1988 og 1996. De fleste prøvene er tatt om høsten, unntaket er prøvene fra 1990, som ble tatt i mars, men de er uansett ikke lett å utnytte i noen analyse av varianskomponenter.

**Tabell 24.** Forekomst av data for krabbesmør fra Kristiansandsfjorden, vist som antall data pr. område og år.

Delområde	1988	1990	1992	1996
B	1	1	3*#	1
D	1&		2*	1
E	1			1
F	1	1&		
G	1&			1

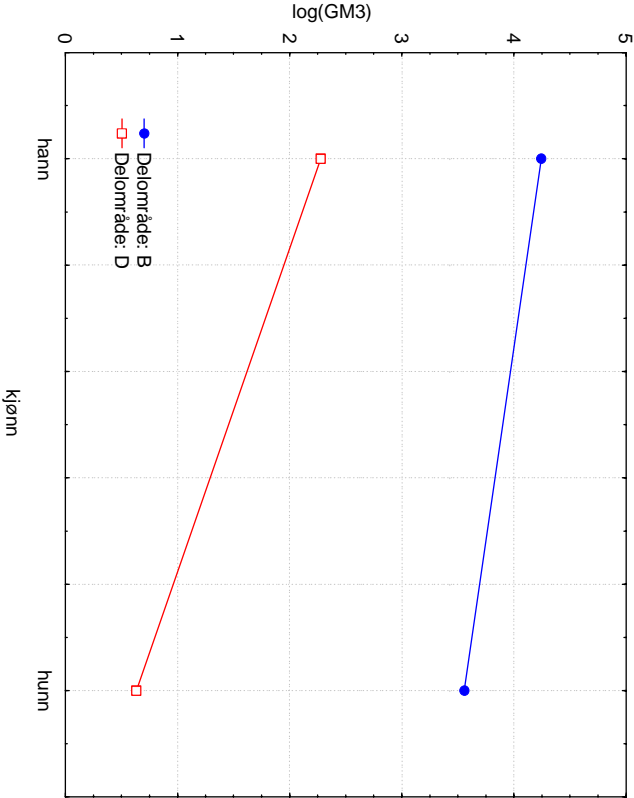
#: En prøve er analysert parallelt av IFFH og NILU – bare NILU-resultatet er tatt med i analysen.

\*: En prøve av humnkrabbe

&: blandet prøve hanner og hunner

8.2.2 Statistiske analyser

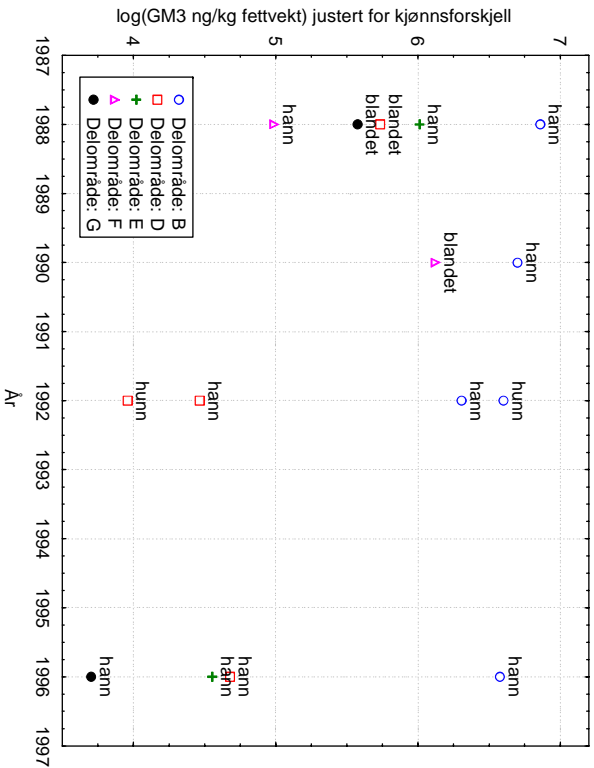
Parallele prøver av hann og humnkrabbe fra område B og D i 1992 undersøkes i en toveis varians-analyse med kjønn og område som faktor og interaksjon kjønn\*område som residualled. Estimerte forskjeller mellom kjønn og område er sterkt avhengig av hva slags variansmodell som blir brukt (faste eller tilfeldige faktorer, interaksjoner uavhengige eller bundet av at summen =0 ) og er ikke i noe tilfelle signifikante. Størrelsen på kjønnseffekten strider ikke mot forskjellen på ca. 0.8 som ble funnet i Grenlandmaterialet, og går samme vei for begge områder og begge kjønn (**Figur 14**). Midlere kvadratavvik for interaksjonen kjønn\*område vil være en sum av varians for utvalgsusikkerhet (mellom replikate prøver) innenfor samme område, tidspunkt og kjønn, og uavhengig variasjon i tid i ulike områder, justert for midlere kjønnsforskjell. Midlere kvadratavvik for interaksjonen estimeres til ca. 0.2 for kombinasjonsmålene på logskala, men basert på bare én friletsgrad, og selvsagt svært usikkert; nedre 10 % konfidensgrense er 0.075, så resultatet strider ikke mot det som ble funnet for Grenland.



**Figur 14.** Forskjell mellom dioksinninnhold i parallelle prøver av hann- og humnkrabbe fra Kristiansandsfjorden 1992.

Prøver av hannkrabbe fra områdene B og E i 1988 og 1996 er studert for om mulig å gi supplerende informasjon om varianskomponenter, men data er for sprikende til at det virker fornuftig. I 1988 ga de to prøvene omtrent like konsentrasjoner. I prøvene fra 1996 er konsentrasjonene noe høyere enn i 1988 fra område B (forskjell ca. 0.8 på logskala), og mye lavere i område E (forskjell ca. -1.5 på logskala). Det gjør at interaksjonen år\*område får en svært høy verdi på 1.26, men bestemt med bare én frhetsgrad.

For å forsøke å få mer ut av datasettet er det gjort en tilleggsanalyse på log(GM3). Først er verdiene justert for kjønnsforskjellen ved å legge til 0.8 for prøvene av hunnkrabbe, basert på den kjønnsforskjellen som ble funnet i Grenlandsprøvene. Prøvene som er blanding av hann og hunn er justert best mulig, dvs. med halvparten av dette. I disse data ser det ut til å være en sammenheng med fett-%. Så er det gjort en variansanalyse med år og delområde som faktorer, interaksjon År\*delområde som del av residualvariansen, og med log(fett-%) som kovariat. Dette gir som resultat at alle effektene er signifikante. Residualvariansen, som i hovedsak er interaksjon år\*delområde, blir ca. 0.21, bestemt med 6 frhetsgrader. Hvis en skiller ut interaksjonen som eget ledd, blir både varianskomponenten for interaksjon og varians mellom replikater estimert til ca. 0.12, med hhv. 0.4 og 1 frhetsgrad, altså svært usikkert. Når det gjelder sammenhengen med fett-% estimeres en koefisient 2.0, med standardavvik omkring 1.0, litt ulikt med og uten interaksjon som eget ledd i variansmodellen. Dette resultatet strider ikke mot at konsentrasjonene i krabbesmør kan være omtrent proporsjonale med fett-%, og samvariasjonen med fett-% kan også være svakere, omtrent slik som estimert for Grenlandsdata . **Figur 15** viser konsentrasjonene regnet om til fett-basis og justert for kjønnsforskjeller. Det er liten eller ingen endring i område B fra 1988 til 1996, men en mye større reduksjon i områdene D, E og G. For område D ser det ut til at reduksjonen har skjedd fra 1988 til 1992, mens det ikke er noen videre reduksjon etterpå.



**Figur 15.** Kjønnsjusterte dioksinkonsentrasjoner på fett-basis (hunnkrabbe justert med +0.8, prøver av blandet kjønn justert med +0.4) som log(GM3 ng/kg fettvekt) for krabbesmør fra Kristiansandsfjorden.

Datamaterialet fra Kristiansandsfjorden er for lite til å gi særlig brukbare estimater for varians eller sammenheng med kjønn eller fett-%, men innenfor de svært store usikkerhetene kan det ikke påvises avvik fra det som ble funnet for Grenland.

8.3 Sammenfattende konklusjon for krabbe

Ut fra tilgjengelige data for parallellprøver av hann- og hunn-krabbe kan det anslås at hannkrabbene på høsten har i overkant av dobbelt så høye dioksinkonsentrasjoner som hunnkrabber under ellers like forhold (forskjell 0.8 i  $\ln(\text{konsentrasjon})$ , tilsvarende faktor 2.23 på konsentrasjonen). Det er også funnet signifikant sammenheng med fett-%, anslått til at lineær regresjon mellom  $\ln(\text{konsentrasjon})$  og  $\ln(\text{fett-\%})$  har en regresjonskoeffisient 0.42, det vil si at konsentrasjonen er estimert å variere proporsjonalt med  $[\text{fett-\%}]^{0.42}$ . Med korreksjon for kjønn og fett-% gir analysen av data fra Grenland brukbare variansestimater både for variasjon mellom replikate prøver innenfor samme stasjon og år, og for reell variasjon i gjennomsnitt mellom ulike kombinasjoner av stasjon og år, dvs. uavhengig variasjon mellom år på hver stasjon. Den gir også et estimat for fluktuasjoner mellom år som gjennomsnitt over stasjoner, men med større usikkerhet. Datamaterialet fra Kristiansandsfjorden er lite og gir bare usikre svært usikre estimater, men det som er funnet strider ikke mot estimatene fra Grenland. Estimater og 80-% konfidensintervaller for enkelte varianskomponenter er oppsummet i

Tabell 25.

Tabell 25. Estimerte varianskomponenter for krabbe

Varianskomponent	Område	Effektivt antall frihetsgrader	Beste estimat	80 % konfidens-intervall
Variasjon mellom replikate prøver fra samme tid og sted	$\sigma_r^2$	Grenland	17	0.081
		Kristiansand	<1	0.12
Reelle uavhengig variasjon mellom år hver stasjon	$\sigma_{st}^2$	Grenland	7	0.074
		Kristiansand	1	0.12
Fluktuasjoner mellom år som gjennomsnitt over stasjoner	$\sigma_t^2$	Grenland	4	0.042
				0.011 til 0.08

Etter en samlet vurdering virker det rimelig å basere vurdering av utsagnskraft på et variansestimater 0.08 for varians mellom replikate prøver fra samme tid og sted og 0.09 for interaksjon sted\*år, etter korreksjon for fett-% og kjønn. Den siste variansen er da justert opp noe for å kompensere for tendensen til underestimering ved få frihetsgrader, se Vedlegg G, og ut fra de litt høyere estimatene fra Kristiansandsfjorden. Om det vil være noen vesentlig varians mellom år som gjennomsnitt over stasjoner er usikkert, og det vil være avhengig av konkret plassering av stasjoner, slik at det antagelig er vanskelig å overføre det direkte til Kristiansandsfjorden, men Grenlands materialet indikerer et variansbidrag på 0.05 hvis estimatet justeres for tendensen til å underestimere (Vedlegg G).

Varianskomponentene gjelder for variasjon etter korreksjon for kjønn og fett-%. Analysen av data fra Grenland viser at begge faktorer er signifikante, med kjønn som den viktigste faktoren. Datasettet fra Kristiansandsfjorden er så lite at det gir svært usikre estimater av sammenheng med kjønn eller fett-%, men estimatene strider ikke mot det som er funnet for Grenland.

Middlere reduksjon over tid er estimert til 9 % pr. år i Grenland i perioden etter de store utslipps-reduksjonene. Datamaterialet fra Kristiansandsfjorden er for lite til å gi brukbart estimat for tidstrend.

De statistiske analysene av tidstrender for dioksininnhold i krabbesmør basert på datamaterialet fra Kristiansandsfjorden og Grenlandsfjordene tilsier:

- Hvis tidsutviklingen i dioksinkonsentrasjoner på ett sted overvåkes ved at en tar blandprøver av krabbesmør fra ca. 20 individer pr. prøve over en årrekke, vil konsentrasjoner for enkelprøvene variere tilfeldig fra år til år rundt jevn tidstrend. Disse variasjonene har relativt standardavvik 40 til 50 % (varians 0.17 for naturlige logaritmer av konsentrasjoner).
- Opp mot halvparten av variansen ser ut til å være utvalgsusikkerhet pga. tilfeldig variasjon mellom individer fra samme sted og tid (varians 0.08 for naturlige logaritmer av konsentrasjon). Resten skyldes antagelig reelle avvik fra trenden i gjennomsnittet fra år til år (varians 0.09 for naturlige logaritmer av konsentrasjon). Hvis det analyseres flere prøver fra samme stasjon hvert år, eller hvis antall individer pr. prøve økes, vil virkningen av utvalgsusikkerheten avta, mens virkningen av de reelle avvikene vil være den samme.
- Hvis det tas parallelle tidsserier fra flere stasjoner innenfor et område, kan en regne med at disse avvikene fra tidstrendene er uavhengige på hver stasjon. Virkningen på samlet års-gjennomsnitt vil da minske jo flere stasjoner en har. I tillegg ser det ut til å være avvik over som er felles for alle stasjoner innenfor et større område (estimert varians 0.05 på naturlig logskala ut fra Grenlandsmaterialet). Et slikt variansbidrag får betydning hvis det samles data fra mer enn 2 stasjoner.
- Den variasjonen som er angitt ovenfor gjelder for konsentrasjoner korrigert for kjønn og fett-%. Kjønn er den viktigste faktoren å korrigere for. Det er viktig å kontrollere og registrere dette ved prøvetaking.
- Data fra Grenland viser ca. 9 % reduksjon av konsentrasjoner i krabbesmør pr. år over en 10-års periode etter store utslippsreduksjoner. Det antas at det er realistisk å vente det samme i Kristiansandsfjorden etter de store utslippsreduksjonene som har skjedd der siden 1990.

# 9    Utsagnskraft ved hypotesetesting som funksjon av omfang og design av overvåkningsprogram

## 9.1    Varianskomponenter fra eksisterende data – en oppsummering

Hoved-resultatene av de forskjellige statistiske analysene i kapitlene foran er sammenstilt i **Tabell 26**. Tabellen viser estimater og beregnede konfidensintervaller for variansbidrag for ulike typer uregelmessig variasjon av logtransformerte verdier av GM3<sup>1</sup>. For blåskjell og torskelever er varianskomponent for interaksjon stasjon\*år ikke estimert direkte i statistiske analyser, men anslått som forskjell mellom sumvariens [interaksjon+residual] og residualvariansen, og da er ikke konfidensintervall er angitt.

**Tabell 26.** Sammenfatning av estimerte varianskomponenter fra Grenland og Kristiansandsfjorden for naturlige logaritmer av kombinert dioksimål GM3, med konfidensintervaller der de er beregnet. df=frihetsgrader for beste estimat av  $\sigma_{st}^2 + \sigma_r^2$

		Estimater og 80 % konfidensintervaller for varianskomponenter				
Prøvetype	Antall individer i prøven	Fluktasjoner mellom år, felles over stasjoner	Fluktasjoner mellom år, uavhengig på hver stasjon	Mellom replikate prøver fra samme stasjon og tid	Samlet varians mellom prøver fra ulike stasjoner samme år	
		$\sigma_i^2$	$\sigma_{st}^2$	$\sigma_r^2$	$\sigma_{st}^2 + \sigma_r^2$	
Blåskjell	50	<0.025	0.05	0.07 (0.03-0.18)	0.11 (0.08-0.16) df=30	
Torskelever	20	<0.02	0.07	0.09 (0.05-0.21)	0.16 (0.13-0.23) df=37	
Skrubbefilet	20				0.14 (0.09-0.31) df= 9	
Krabbesmør	20	0.05 (0.011-0.08)	0.08 (0.05-0.12)	0.09 (0.03-0.12)	0.17 (0.13-0.23) df=49	

Tabellen viser at resultatene for de forskjellige prøvetypene er svært like når usikkerheten i estimatene tas i betraktning. For skrubbeflyndre er det ikke grunnlag for å dele opp variansen, men totalvariansen for interaksjon år\*stasjon er av samme størrelse som for de andre prøvetypene. Forholdstall mellom variansestimater for forskjellige prøvetyper varierer omtrent som en vil vente for forskjellige estimater for samme varians<sup>2</sup>. Det er derfor ikke noe sikkert grunnlag for å peke ut noen av prøvetypene som mer egnet enn andre til å detektere trender i dioksinkonsentrasjonene, i alle fall ikke når en begrenser vurderingene til de mest detekterbare dioksinkomponentene.

<sup>1</sup> GM3 er geometrisk middelværdi av 2378-TCDF, 12378/12348-PeCDF og 23478-PeCDF (se kapittel 3.6)

<sup>2</sup> Hvis en har to uavhengige estimater for samme varians  $s_a^2$  og  $s_b^2$  med frihetsgrader  $v_a$  og  $v_b$ , vil forholdstallet mellom dem være F-fordelt med ( $v_a$ ,  $v_b$ ) frihetsgrader.

Resultatene kan riktignok peke i retning av at blåskjellprøvene har noe mindre fluktusjoner enn de andre prøvetyperne. Dersom det er tilfelle, kan det skyldes både mindre variabilitet i populasjonen på stor og liten skala og mindre utvalgusikkerhet pga. de prøvestørrelsene som er brukt. Forskjellen er likevel marginal i forhold til usikkerheten i estimatene, og valg av organismer for videre overvåking må derfor vurderes ut fra andre hensyn, f.eks. betydning for kostholdsråd. Når formålet er å sammenligne med tidligere data er det dessuten viktig å kunne bruke så mye som mulig av det tidligere innsamlede materialet, spesielt siden det finnes så lite data, så det kan være en grunn til å ta nye prøver av alle de prøvetyperne som det er tidligere data for i noe omfang.

Oppdelingen i ulike varianskomponenter er usikker. Imidlertid gir de uavhengige analysene på ulike prøvetyper nok så like resultater, så det kan være grunn til å tro at oppdelingen er brukbar for å skissere utsagnskraft. Som grunnlag for analyse av utsagnskraft og overvåkningsprogram generelt for alle prøvetyper brukes derfor følgende verdier for varians av log(konsentrasjon) i forhold til gjennomsnittlige stasjonsnivåer og jevn tidsutvikling, som et gjennomsnitt av de forskjellige resultatene:

	Varians	Standardavvik
Fluktusjoner mellom år, felles over stasjoner	$\sigma_t^2$	0.02
		0.14
Fluktusjoner mellom år, uavhengig på hver stasjon	$\sigma_{st}^2$	0.07
		0.26
Mellom replikate prøver fra samme stasjon og tid	$\sigma_r^2$	0.09
		0.30

Variansen mellom replikater kan reduseres ved å ta med flere individer i blandprøvene. Sammenhengen mellom antall individer og varians for logtransformerte verdier i blandprøver er analysert i Vedlegg H.

Estimatene angir varianskomponenter som bidrar til variasjon mellom enkeltprøver eller gjennomsnitt etter følgende skjema:

Variasjon mellom enkeltprøver:	har total varians:	totalt standardavvik
– fra samme tid og sted	$\sigma_r^2$	= 0.09
		0.30
– fra forskjellige stasjoner samme år	$\sigma_{st}^2 + \sigma_r^2$	= 0.16
		0.40
– fra forskjellige år, enten det er samme stasjon eller fra forskjellige stasjoner	$\sigma_t^2 + \sigma_{st}^2 + \sigma_r^2$	= 0.18
		0.42
Variasjon mellom årsgjennomsnitt over $S$ stasjoner samme år med $R$ replikater pr. stasjon i forhold til jevn trend:	$\sigma_t^2 + \frac{1}{S}\sigma_{st}^2 + \frac{1}{SR}\sigma_r^2$	

Ut fra de konfidensintervallene som er satt opp ovenfor for variansbidragene kan det være rimelig å hevde med 80 % konfidens at relativ usikkerhet i anslaget for standardavvik ligger innenfor 20 %, iallfall for prøvetyperne blåskjell, torskelever og krabbesmør.

I Vedlegg B er det satt opp variansuttrykk for fluktusjoner av årsmidler og enkeltprøver fra ulike kombinasjoner av delområde og år når hver blandprøve er basert på flere taking fra flere lokaltiteter.

## 9.2 Vurdering av utsagnskraft ved statistiske tester av data fra ny overvåking

I det følgende brukes variansestimaterne til å vurdere utsagnskraft ved ulike prøvetakingsprogrammer i forhold til noen konkrete problemstillinger.

Variansestimaterne er i hovedsak basert på data fra Grenlandsområdet. Estimaterne fra Kristiansands-fjorden er svært usikre fordi de bygger på så lite data, men det er ikke funnet noen tilfeller av markerte avvik mellom de to områdene. Det kan derfor være forsvarlig å anta at estimatene fra Grenland kan brukes for Kristiansandsfjorden.

Det er rimelig å tro at variansen mellom replikate prøver tatt på samme tid og sted vil være nokså generell og overførbar mellom områder. De to andre varianskomponentene, for variasjon mellom år og mellom kombinasjoner av år og stasjon, kan være mer avhengig av konkrete forhold i hvert område. Forskjellige områder kan ha ulik grad av påvirkning fra varierende restutslipp i forhold til sekundær belastning fra sedimenter, og ulik grad av variasjon mht. spredning/mobilisering av forurensning. Det kan også være forskjeller i hvordan valg av tidspunkter og steder for prøvetaking slår ut i forhold til variasjonsmønstre, som det ikke grunnlag for å kartlegge.

Varianskomponentene gir uansett bare en forenklet beskrivelse av en sammensatt variasjon på mange rom- og tids-skalaer. Vi har ikke grunnlag for å vurdere dette nærmere, og må bare gå ut fra at de estimerte varianskomponentene kan brukes til å vurdere varians for et prøvetakingsprogram hvor det tas prøver fra stasjoner som har stor grad av uavhengige flukturasjoner og at den estimerte interaksjonen år\*stasjon er omtrent riktig. Hvis stasjonene er nær hverandre, vil variansen år\*stasjon bli mindre, og prøvene fra forskjellige stasjoner får mer karakter av replikatprøver fra samme sted. Også hvis prøver fra samme stasjon tas flere ganger pr. år, f.eks. med 1 til 2 måneders mellomrom kan det være at de bør betraktes som replikatprøver.

Variansestimaterne antyder at bidraget fra variasjon på litt større skala, dvs. mellom ulike steder og år, omtrent den samme som variasjon mellom replikater tatt på samme tid og sted. Den totale variansen mellom enkeltprøver fra ulike stasjoner eller år er en sum av de to variansbidragene. Hvis variansen mellom replikater skyldes ren utvalgsusikkerhet, dvs. variasjon mellom individer, vil en dobling av individtallet pr. prøve redusere denne variansen med en faktor 0.6 til 0.55 på logtransformerte verdier (Vedlegg H). Den totale variansen mellom prøvene i en tidsserie i forhold til trenden vil også omfatte variasjonen mellom steder og år og reduseres derfor mindre, med en faktor på ca. 0.8 på varians, eller 0.9 på standardavvik. Dvs. at evnen til å oppdage endringer kan øke med ca. 10 %, målt som den ending som kan oppdages med gitt teststyrke og signifikansnivå. Hvis en fordeler innsamlingen av samme mengde materiale til en blandprøve fra flere tider og steder gjennom året, vil en også få redusert en del av den variansen som skyldes variasjoner på mindre skala i tid eller rom og kan øke teststyrken vesentlig mer for omtrent samme kostnad. Det er usikkert datagrunnlag for å gjøre slike vurderinger, fordi det er så få replikate prøver, og ikke detaljerte beskrivelse av hvordan prøvene er tatt mht geografisk utstrekning eller utstrekning i tid, men det gjøres likevel noen vurderinger av dette i det følgende.

I det følgende skal vi se på tre ulike typer statistiske tester og vurdere utsagnskraft som funksjon av omfang av nytt overvåkningsprogram. Vi ser da bare på test på data for en prøvetype av gangen. Det kan også være aktuelt å se resultatene av testene for ulike prøvetyper samlet og se om de bekrefter hverandre og gir grunnlag for sterkere konklusjoner, slik det for eksempel er gjort av Bjerkeng og Ruus (2002), men det vurderes bare helt summarisk. Vurderingen av utsagnskraft gjøres ut fra en forutsetning om at testene baseres på variansestimater ut fra de data som inngår i testen, og at det ikke inngår variansestimater fra andre lignende data.



### 9.2.1 Utsagnskraft for tester på om dioksin-nivåene i organismer redusert siden 1996

Et viktig mål med ny overvåking i Kristiansandsfjorden vil være å se om det har skjedd en nedgang i nivåene siden 1996 vurdert ut fra nye dioksimålinger i organismer i 2006-2007. Det er allerede målt vesentlig lavere konsentrasjoner i blåsjeell fra 2002 (**Figur 4**) sammenlignet med 1996, etter to år med spesielt lave utslipp i 2001 og 2002. Utslippene var høyere igjen i 2003 og 2004, men skal nå være kommet ned på samme lave nivå som i 2001 og 2002 (Se kapittel 1). Spørsmålet er hvor stor en eventuell endring fra 1996-nivåene må være for at den skal kunne påvises noenlunde sikkert avhengig av omfang og design av ny prøvetaking. I denne sammenhengen ser vi på hva som kan oppnås ved ny overvåking i 1-3 år.

Når data fra to enkelt-år sammenlignes med hverandre, vil variasjon som på lengre sikt fremtrer som fluktusjoner inngå i den effekten som påvises og ikke som feilledd. Det vil ikke være mulig å skille mellom forskjell som skyldes uregelmessige utslag på kort sikt, f.eks. variasjon i utslipp som beskrevet ovenfor, og det som skyldes mer varige endringer.

- Hvis en sammenligner gjennomsnitt over flere stasjoner, vil tilfeldig variasjon fra år til år som er felles for alle stasjoner (varians  $\sigma_r^2$ ) inngå i den effekten som skal testes. Analyse av data fra Grenland indikerer at dette variansbidraget er innenfor 0.04 på logskala. For års-gjennomsnitt over 4 stasjoner indikerer variansestimaten i kapittel 9.1 at de uavhengige variasjonene på hver stasjon  $(\sigma_{st}^2 + \sigma_r^2)/4$  bidrar like mye til fluktusjon i årsgjennomsnittene som den variasjonen som er felles for stasjonene. Den totale variansen for fluktusjoner i åtsgjennomsnittene blir da 0.08 på logskala. Forskjellen mellom logtransformerte gjennomsnitt for to adskilte enkeltår i prøvematerialet vil da variere rundt den 'sanne' verdien (for hele populasjonen/hele området) med en varians på det dobbelte av utslaget for et enkelt år, dvs. 0.16. Regnet om til forhold mellom laveste og høyeste årsmiddel (geometrisk middel) blir det  $0.67 = \exp(-\sqrt{0.16})$ . Hvis det kan påvises vesentlig større reduksjoner enn det, kan de være del av en varig forbedring.

- Hvis en sammenligner data for en og en stasjon, vil samlet variasjon mellom enkeltprøver i forhold til generelt gjennomsnitt ( $\sigma_r^2 + \sigma_{st}^2$ ) inngå bare i effekten. Analysen av data antyder at slik fluktusjon mellom år for enkeltstasjoner har varians 0.11 på logskala, tilsvarende en forskjell med en faktor  $\exp(-\sqrt{0.22})$  for eksempel som reduksjon med en faktor 0.62 på lineær skala. Hvis det er vesentlig større forskjell enn dette, kan de være del av en varig endring. En test på forskjell mellom år for enkeltstasjoner forutsetter at det nye programmet gjennomføres med replikate prøver fra samme stasjon og periode, slik at en har et grunnlag for å estimere variansbidraget  $\sigma_r^2$ .

Hvis overvåkingen fremover utvides til å gå over noen år, kan det være mulig å se om det er signifikante forskjeller ut over det som kan være resultat av fluktusjoner fra år til år. Testen vil uansett bare gå ut på å vise om dagens nivåer er stabilt lavere enn nivåene i 1996. Hvor representative nivåene målt i 1996 er for utviklingen etter 1992 kan bare anslås ved å anta at relative fluktusjoner i årene omkring 1996 er omtrent de samme som i dag, estimert på grunnlag av nye data. En slik antagelse vil ligge innbakt i de testene som gjøres. Det er for lite data fra Kristiansandsfjorden for perioden etter utslippsreduksjonene i 1992 til at en nå kan estimere direkte hva variasjonsmønsteret vil være i en periode hvor sekundære, forsinkede virkninger av dioksin lagret i sedimenter er viktigere enn før i forhold til fortsatte utslipp. Estimaten fra Grenland, som også er relativt usikre, kan ha begrenset overførbarhet til andre områder og måleprogrammer. Hvis dagens utslipp har vesentlig betydning for nivåene, og de varierer slik de har gjort i perioden 2000-2004, kan det uansett være vanskelig å si hva en forskjell fra 1996-nivå representerer, og da kan det tenkes at det vil være nødvendig å ta med variasjon fra år til år som er felles over arter og/eller stasjoner som faktor i de statistiske analyser.

Med disse forbeholdene kan det anslås hva en vil kunne påvise av endringer fra 1996 til 2006-2007 ved ny prøvetaking. Både for blåskjell, torsk, skrubb og krabbe finnes data fra 1996 som sammen-igningsgrunnlag, og for blåskjell også fra 1995. Sammenligningsgrunnlaget er lite, og det begrenser hvor små endringer som kan oppdages selv om en tar ganske mange prøver i et nytt program. Usågs-raften i et overvåkningsprogram vil avhenge av hvilke antagelser en finner å kunne gjøre for de forskjellige variansbidragene som estimeres i en variansanalyse.

I denne sammenhengen ser en altså på forskjell mellom de konkrete årene det er data fra, og tid inngår derfor som fast faktor, men sted kan betraktes på ulike måter:

I. Hvis også prøvetakingsstedene er valgt ut som steder av konkret interesse og en bare vil si noe om gjennomsnittlig tidsutvikling på de utvalgte stedene, skal *sted* behandles som såkalt fast faktor. Da er interaksjonen definert i forhold til de konkrete stedene og prøvetidspunktene, og inngår ikke i hovedeffektene. Både hovedeffekter og interaksjon testes mot residualvariens, som antas å være det eneste "tilfeldige" elementet i datamaterialet. Konklusjonene gjelder det faktiske utvalget av prøvesteder og ikke som generelle konklusjoner om området. For å kunne gjøre dette må det foreligge replikate prøver for i alle fall en del av kombinasjonene sted\*år.

II. Hvis prøvetakingsstedene betraktes som tilfeldig valgt og en vil si noe om området som helhet, dvs. om et gjennomsnitt også over andre mulige prøvesteder, skal *sted* inngå som tilfeldig faktor. Da vil interaksjonen være et tilfeldig uttrykk for uavhengig variasjon i tid mellom steder, som også gir en tilfeldig restvirkning i de estimerte hovedeffektene, f.eks. gjennomsnittet over stasjoner pr. tidspunkt. Hovedeffektene testes mot interaksjonen, for å se om de er noe mer enn slike tilfeldige restvirkninger av fluktuasjoner. Uten replikater er det slik testen må gjøres.

Alternativ I vil gi større sannsynlighet for å påvise endringer i tid i gjennomsnitt over de valgte prøvestedene, uten hensyn til om det er representativt for området som helhet. Alternativ II innebærer en vurdering av hvor tydelig endringen er i forhold til det som kan skyldes tilfeldige utslag på de ulike prøvestedene. Det antas at alternativ II er mest relevant for en vurdering av miljøstatus, og i beregningene som følger er det stort sett antatt at testene vil bli gjort på den måten. Det er også mest realistisk pga. mangel på replikater.

En variansanalyse vil estimere midlere kvadrat for de forskjellige effekter, og teste for signifikans ved å sammenligne midlere kvadrat i F-tester. For et balansert datasett med *N* tidspunkter (f.eks. år eller perioder), *S* steder (stasjoner eller delområder) og *R* replikater pr. kombinasjon Tid\*Sted er forventningsverdiene for kvadratavvikene lineære kombinasjoner av de forskjellige varians-komponentene, som vist nedenfor:

Effekt	Varians-komponent	Forventet midlere kvadrat (expected mean square EMS)		
		Frihets-grader	Forventningsverdi med sted som fast faktor	Forventningsverdi med sted som tilfeldig faktor
Tid	$\sigma_t^2$	N-1	$NR\sigma_t^2 + \sigma_r^2$	$NR\sigma_t^2 + R\sigma_{st}^2 + \sigma_r^2$
Sted	$\sigma_s^2$	S-1	$SR\sigma_s^2 + \sigma_r^2$	$SR\sigma_s^2 + R\sigma_{st}^2 + \sigma_r^2$
Interaksjon Tid*Sted	$\sigma_{st}^2$	(N-1)(S-1)		$R\sigma_{st}^2 + \sigma_r^2$
Replikater	$\sigma_r^2$	NS(R-1)		$\sigma_r^2$

Ubalanserte datasett hvor antall replikater varierer, eller hvor ikke alle kombinasjoner av tid og sted har data, vil ha en lignende oppdeling, men med andre faktorer, og de ulike leddene er ikke helt uavhengig av hverandre, slik at testene er mindre veldefinerte.

Hvis det ikke er replikater ( $R=1$ ), må sted betraktes som tilfeldig faktor og hovedeffekter testes mot interaksjon. Hvis det er replikater, men ingen signifikant interaksjon, kan MS for interaksjon og replikater eventuelt slås sammen og brukes som feilledd ved test av hovedeffekter.

Hvis det er grupper av år eller steder som ikke er signifikant forskjellig innbyrdes, kan modellen eventuelt forenkles ved at prøver innenfor gruppene betraktes som replikater innenfor en grovere oppdeling. Hvis en har data fra flere stasjoner eller områder, men det ikke er tegn til signifikante geografiske forskjeller mellom stasjonene som gjennomsnitt over tid, kan en eventuelt forenkle til en enveis variansanalyse med tid som faktor, hvor alle prøvene fra samme tid betraktes som replikater. For å vurdere dette er det naturlig å ta utgangspunkt i total varians for interaksjon stasjon\*år uten replikater, som er estimert til 0,16 for logtransformerte verdier, dvs. et standardavvik 0,4 på logskala, eller en faktor 1,5 på konsentrasjon. Forskjeller mellom stasjoner i gjennomsnitt over tid som er vesentlig større enn dette, må en regne med kan være mer varige, slik at stasjon må med som faktor i analysen.

Det å se bort fra ikke signifikante effekter ved å forenkle modellen vil kunne bety at teststyrken øker mht. å påvise effekter, men at signifikansnivået (risikoen for å påvise effekter der det ikke er noen) også blir høyere. Hvis antagelsene holder stikk, vil innledertid resultatet bli riktigere. Det vil være en vurderingssak om det skal forenkles; det mest konservative vil være å beholde alle definerte faktorer selv om de ikke er signifikante. Nedenfor vurderes teststyrke ut fra ulike vurderinger der det er naturlig.

Hvis vi bare skal se på forskjell mellom to perioder ( $N=2$ ), brukes en F-test som sammenligner varians mellom periodene med varians mellom prøver innenfor periode. Dette er ekvivalent med en t-test hvor forskjellen mellom periodene testes mot standardavvik mellom prøvene innenfor periode.

I de følgende avsnittene er det vurdert hva en kan regne med å oppnå ved ny overvåking av de ulike prøvetypene for sammenligning med data fra 1996. Vurderingene er gjort separat for hver prøvetype. Når nye resultater foreligger kan det være aktuelt å gjøre en samlet vurdering hvor data fra de ulike artene ses i sammenheng, men det er sett på dette her – det må vurderes ut fra resultatene om og på hvilken måte det eventuelt skal gjøres.

9.2.1.1 Blåskjell

For blåskjell er data fra område A og B i 1995 og 1996 mest relevante for en sammenligning. De fleste av dataene fra 1995 og 1996 er fra JAMP-programmet, og det vil derfor være en fordel å benytte de samme prøvestedene i en ny overvåking, dvs. Fiskåtangen og Odderøy Vest, og også fra Silokaia hvor det ble tatt en prøve i 1996. Det har også verdi å ta nye prøver fra Myrodden, hvor det ble tatt prøve i 1990 samtidig med en prøve fra Silokaia, for å få best mulig bilde av varians mellom prøvestedene, og eventuelt utnytte det til å forbedre en senere analyse.

Hvis det tas prøver i 2006 og eventuelt i 2007 fra de stedene det ble tatt prøver i 1995 og 1996, vil vi ha følgende sett av prøver:

		Område		A		B	
		Stasjon	Myrodden	Fiskåtangen	Sliokaia	Odderøy vest	
Periode		År					
Før		1995		1			1
		1996		1	1		1
Nå		2006	1	1	1		1
		2007	1	1	1		1

En kovariansanalyse av  $\log(\text{GM3})$  fra 1995, 1996 og 2002 med delområde som kategorisk faktor og år som regresjonsvariabel (dvs. antatt lik relativ variasjon over tid) viser en gjennomsnittlig forskjell 0.3 på logskala mellom delområdene. Forskjellen er signifikant i forhold til residualvarians 0.065 (standardavvik 0.25) for enkeltprøver i det utvalgte datasettet, men ikke i forhold til det generelle estimatet for total varians i interaksjon stasjon\*år på 0.16 (standardavvik 0.4).

Det kan derfor være at det blir aktuelt å betrakte alle prøver innen periode som replikater, uten hensyn til geografiske forskjeller, og evt. bruke også nye data fra Myrodden. Det kan også være at nye resultater indikerer signifikant forskjell mellom stasjonene som gjennomsnitt over tid. Det vil da være nødvendig å ta med stasjon som faktor i analysen, og bare bruke data fra de stasjonene hvor det ble tatt prøver i 1995.

### Teststyrke hvis det ikke er forskjell mellom områder eller stasjoner.

Hvis forskjellen mellom år innenfor hver periode som gjennomsnitt over stasjoner er liten sammenlignet med forskjellen mellom periodene, kan data fra 1995-96 ses samlet som replikater fra periode  $F\phi r$  og alle data fra 2006-2007 som replikater fra periode  $N\hat{a}$ . Testen kan isåfall gjøres som en enkel t-test av forskjellen mellom gjennomsnitt av  $\log(C)$  for før- og nå-perioden, hvor  $n_1=5$  prøver fra  $F\phi r$ -perioden sammenlignes med  $n_2 =$  hhv. 4 og 8 prøver fra ny overvåking over ett eller to år, i  $N\hat{a}$ -perioden med test-verdi

$$t = \frac{\overline{\log(C)_{F\phi r}} - \overline{\log(C)_{N\hat{a}}}}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

med  $v = n_1 + n_2 - 2$  og frihetsgrader og hvor  $s$  er kvadratroten av estimatet for varians  $\sigma$  for enkeltprøver.

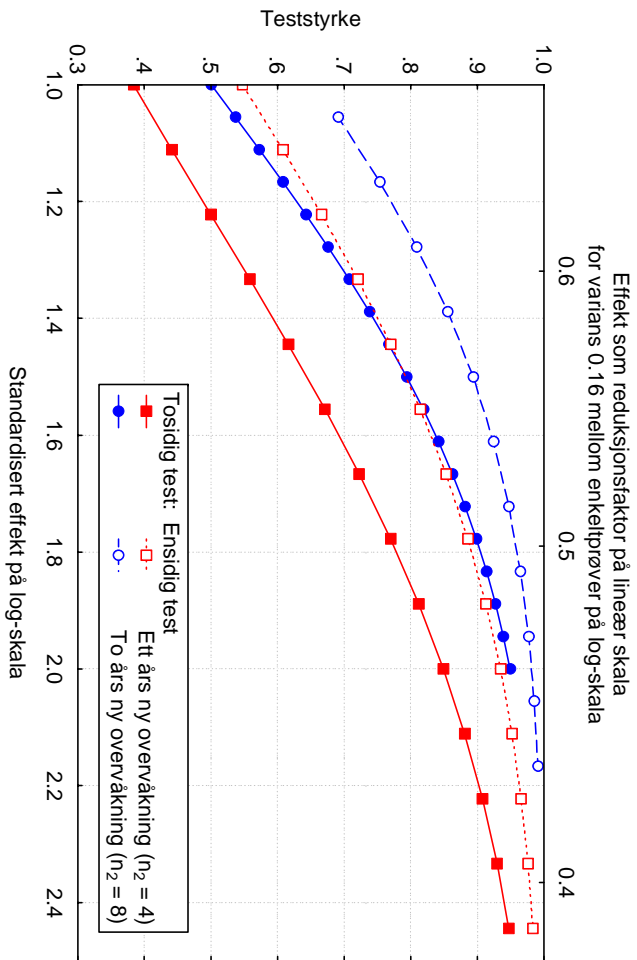
Estimert utsagnskraft for slike tester basert på  $s = \sqrt{0.16} = 0.4$ , er vist i **Figur 16**. For tosidig test med signifikansnivå  $\alpha=0.1$  må forskjell i forventningsverdi mellom periodene for ny overvåking i ett eller to år være hhv. 1.5 og 1.85 ganger standardavviket for residualer for enkeltprøver for at endringen skal kunne påvises med sannsynlighet (teststyrke) 80 %. Med standardavvik for  $\log(C)$  i enkeltprøver på 0.4 vil det si ca. 0.6 til 0.7 på logskala. Regnet tilbake til lineær skala vil det si at en reduksjon på 45-50 % eller mer kan påvises med 80 % sikkerhet. For å ha 90 % teststyrke må reduksjonen være 50-60 %.

Hvis en økning utelukkes å priori og det brukes en ensidig test med  $\alpha=0.1$ , øker evnen til å oppdage reduksjoner som vist i **Figur 16**; da vil reduksjoner på 40-45 % påvises med 80 % teststyrke og reduksjoner på 45-50 % påvises med 90 % teststyrke.

På grunn av usikkerhet i utgangsverdiene i 1996 vil reduksjoner på rundt 30-40 % ikke kunne påvises sikkert uansett hvor mange nye prøver som tas, hvis en vil ta i betraktning at konsentrasjonene også kan tenkes å ha økt.

En kan oppsummere med at det kan være mulig å påvise endringer hvis konsentrasjonene minst er halvert over 10 år, dvs. hvis relativ reduksjonsrater er på minst 3 % pr år. Forutsetningen for dette er at det ikke er noen signifikant forskjell mellom stasjonene i område A og B i gjennomsnitt over de to periodene.

Sammenligning av to gjennomsnitt med signifikansnivå 0.1 og  $n_1=5$



**Figur 16.** Teststyrke for å oppdage reduksjon i blåskjellkonsentrasjoner i forhold til 1995-1996 ( $n_1=5$  prøver) ved ett til to års ny overvåkning ( $n_2$  prøver) forutsatt ingen systematisk forskjell mellom område A og B og uavhengige avvik for enkelprøver fra gjennomsnitt innen periode, for tester med signifikansnivå  $\alpha=0.1$ .

### Teststyrke hvis det er forskjell mellom stasjoner

Vi antar fortsatt at variansen er 0.16 for enkelprøver i forhold til hovedeffekter Periode og stasjon, og antar at det ikke er noen signifikant forskjell mellom år innen samme periode som gjennomsnitt over stasjoner. Hvis det på hver av de tre stasjonene som har data fra 1995-1996 tas nye prøver i 2006 og eventuelt også 2007, har vi en toveis ANOVA design med antall prøver som vist nedenfor.

Antall prøver pr. kombinasjon av periode og stasjon				
Periode:	Med data fra år:	Fiskåangen	Silokaia	Odderøy vest
Før	1995 og 1996	2	1	2
Nå	2006 (2006 og 2007)	1 (2)	1 (2)	1 (2)

Det antas at data blir analysert i en ANOVA-modell hvor Periode og Stasjon inngår som hoved-faktorer. Prøvene fra ulike år innenfor periode regnes som replikater innenfor Stasjon\*Periode, i det vi antar at reduksjonen fra ett år til det neste som del av en langsiktig trend er mye mindre enn de uregelmessige fluktuationene rundt trenden. Stasjoner betraktes som tilfeldig valgte prøvesteder innenfor det kombinerte område A+B, og det bygges inn i modellen at det kan være konstant forskjell i konsentrasjon mellom stasjonene på logskala, men antas at det ikke er ulik tidsutvikling mellom periodene på de forskjellige stasjonene..

Test på forskjellen i gjennomsnitt mellom perioder i en toveis variansanalyse uten replikater blir en F-test med frihetsgrader ( $v_1=1, v_2 \geq 1$ ). Testverdien for F-testen er forholdet mellom effekt-varians og residualvarians, i dette tilfellet med forventningsverdier ( $EMS = \text{expected mean squares}$ ):

$$EMS_{effekt} = k_F \sigma_{periode}^2 + \sigma^2 ; \quad EMS_{residual} = \sigma^2 ; \quad \sigma_{periode}^2 = \frac{(\overline{\log(C)_{Før}} - \overline{\log(C)_{Nå}})^2}{2}$$

Siden  $v_1=1$ , blir dette ekvivalent med en t-test med  $v_2$  frihetsgrader hvor testverdien er kvadratroten av F-verdien. Testverdien for t-testen vil være

$$t = k_i \frac{\overline{\log(C)_{Før}} - \overline{\log(C)_{Nå}}}{s} \quad \text{med } k_i = \sqrt{\frac{k_F}{2}}$$

på samme måte som over, men med  $k$  definert av hvordan data fordeler seg på de forskjellige kombinasjonene av stasjon og periode.

Siden datasettet er litt ubalansert, er Statistica 7.1 brukt til å finne verdier for  $k_F$  for dette tilfellet. I **Tabell 27** er faktor  $k$  og frihetsgrader  $v$  satt opp for de ulike alternativer mht. overvåkingsomfang og analysemåte, og det er beregnet hvor stor reduksjonen siden 1996 må være for at en reduksjon skal fastslås med 80 og 90 % sannsynlighet. Reduksjonsfaktor for teststyrke 90 % i de to første kolonnene i tabellen svarer til punkter på kurvene i **Figur 16**.

**Tabell 27.** Evne til å påvise reduksjon i dioksinkonsentrasjon siden 1996 i blåskjell i Kristiansandsfjorden som funksjon av statistisk modell og omfang og lengde på nytt overvåkingsprogram ved ensidig test for reduksjon med signifikansnivå  $\alpha=0.1$ . Basert på sammenligning av  $n_1=5$  prøver fra periode 1 (år 1996) med  $n_2$  prøver fra periode 2 (Nå). Det er forutsatt at avvik fra gjennomsnitt for periode (Modell I), eller for stasjon og periode (Modell II), er uavhengige for enkeltprøver.

	Modell I. Uten stasjonseffekt - enkeltprøver som replikater innen periode			Modell II. Med stasjonseffekt - enkeltprøver som replikater innen stasjon*periode		
<b>Antall år med ny overvåking:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>1</b>	<b>2</b>	<b>3</b>
$n_2$	4	8	12	3	6	9
$k_F$	4.44	6.15	7.06	4.33	5.33	6.33
$v$	7	11	15	4	7	10
$k_i$	1.49	1.75	1.88	1.47	1.63	1.78
<b>Reduksjon som kan påvises med teststyrke 90 %</b>	52.1 %	45.6 %	43.0 %	54.7 %	49.0 %	45.3 %
<b>Reduksjon som kan påvises med teststyrke 80 %</b>	45.8 %	39.7 %	37.3 %	48.2 %	42.8 %	39.4 %

Analyse av utsagnskraft for test på om nivåer i blåskjell er redusert i Kristiansandsfjorden siden 1996 viser:

- Dersom konsentrasjonene i blåskjell har blitt redusert med 50 % eller mer siden 1996, vil ett års ny overvåkning på 3-4 stasjoner antagelig kunne påvises reduksjonen med 80-90 % sannsynlighet ved en ensidig test hvor signifikansnivået settes til 10 %.
- Hvis ny overvåkning forlenges til to eller tre år, kan en regne med å påvise reduksjon med samme sikkerhet også hvis det har vært en noe svakere nedgang på 40-45 %.
- Hvis det er klare tegn til vedvarende forskjell mellom stasjonene, kan evnen til å oppdage endringer bli litt mindre, men forskjellen er nok så uvesentlig.
- Overvåkning over 3 år eller mer kan komme til å vise at fluktusjoner fra år til år felles over stasjoner må tas med, og det kan også redusere evnen til å detektere forskjell mellom nivåene i 1996 og et gjennomsnitt over ny overvåkingsperiode. Dette må vurderes når nye resultater foreligger.
- Det kan være realistisk å regne med reduksjonsrater på 10 % pr. år, og det betyr en reduksjon med 65 % i løpet av 10 år. Det er derfor absolutt mulig at en vil kunne påvise signifikant ending allerede etter ett års ny overvåkning av blåskjell, forutsatt at en tar prøver fra 3 eller 4 stasjoner.
- For å få et bilde av variansen i dagens situasjon bør det tas sikte på iallfall tre års overvåkning, men det er altså mulig at en kan trekke foreløpige konklusjoner allerede etter det første året.

### 9.2.1.2 Torskelever

For torskelever finnes data fra 1996 for to prøver fra område B og en prøve fra hvert av de andre områdene. Når eldre data tas i betraktning, er det klar systematisk og varig forskjell mellom områdene, så område må inngå som faktor i en variansanalyse. Område B hadde 3 ganger referansekonsentrasjon i 1996 mens område D og E lå likt omtrent på referansekonsentrasjonen. Forskjellen mellom B og D-E er ca. 1 på logskala, og det er klart større enn hva som skulle ventes tilfeldig ut fra total varians for interaksjon år\*område for torskelever (øvre 80 % konfidensgrense 0.23 for varians, dvs. standardavvik 0.5 på logskala). Ut fra de få prøvene som finnes ser det også ut til at område B og D kan ha hatt omtrent samme reduksjon i gjennomsnitt fra 1988 til 1996, ca. 0.5 på logskala, dvs. ca. 40 % redusert konsentrasjon (**Figur 9**), selv om det er stor usikkerhet forbundet med et slikt anslag for reduksjon. Ut fra de foreliggende prøvene hadde områdene G og F lavere konsentrasjoner enn B, D og E, og område E og F viser ingen reduksjon fra 1992 til 1996. Prøvene fra område E viser faktisk en økning fra 1992 til 1996, men det ligger innenfor det som kan være en tilfeldighet, så det er ingen indikasjon på en generell økning for området. Siden det stort sett ikke finnes replikate prøver, er dette selvsagt bare usikre indikasjoner på forskjeller og tidsutvikling.

For å vurdere evnen til å oppdage endringer over tid ved sammenligning av nye data med tall fra 1996 er det mest aktuelt å se på en analyse av data fra område B (2 prøver i 1996) og område D (1 prøve fra 1996). Det antas at ny prøvetaking tas innenfor 3 år, og at ending over tid i løpet av ny periode er liten i forhold til fluktusjoner mellom enkelprøver. Det forutsettes at alle områdene har omtrent lik relativ reduksjon og at enkelprøver fra forskjellige områder og/eller tidspunkt (år) har uavhengige avvik fra gjennomsnittet for område og periode med varians 0.16 på naturlig logskala. Da kan det gjøre en toveis variansanalyse på område og periode uten interaksjonsledd på samme måte som ovenfor i kapittel 9.2.1.1. **Tabell 28** viser evnen til å oppdage endringer med en slik analyse dersom forutsetningene er oppfylt.

**Tabell 28.** Evne til å påvise reduksjon i dioksinkonsentrasjon siden 1996 i torskelerver i Kristiansandsfjorden som funksjon av omfang og lengde på nytt overvåkningsprogram ved ensidig test for reduksjon med signifikansnivå  $\alpha=0.1$ . Basert på sammenligning av  $n_1=3$  prøver fra periode 1 (år 1996) med  $n_2$  prøver fra periode 2 (Nå). Det er forutsatt at avvik fra gjennomsnitt for område og periode er uavhengige for enkelprøver.

Varighet av ny overvåkning:	Ett år		To år		Tre år	
	1+1	2+1	2+2	4+2	3+3	6+3
<b>Antall nye prøver fra område B+D</b>						
$n_2$	2	3	4	6	6	9
$k_F$	2.33	3.00	3.33	4.00	3.90	4.50
$v$	2	3	4	6	6	9
$k_I$	1.08	1.22	1.29	1.41	1.40	1.50
<b>Reduksjon som kan påvises med teststyrke 90 %</b>	72.1 %	63.3 %	59.5 %	54.5 %	54.9 %	51.3 %
<b>Reduksjon som kan påvises med teststyrke 80 %</b>	65.0 %	56.5 %	52.8 %	48.0 %	48.4 %	45.0 %

Analyse av utsagnskraft for test på om nivåer i torskelerver er redusert siden 1996 viser:

- Etter ett års ny overvåkning av torskelerver med analyse av en blandprøve fra hvert av områdene B og D kan en regne med nokså sikkert å påvise en reduksjon fra 1996 dersom gjennomsnittskonsentrasjonene er redusert med minst 60-70 %.
- Fortsetter overvåkingen i ett eller to år til, vil det være mulig å påvise en reduksjon på 45-55 % med samme sikkerhet.
- En reduksjonsrate ca. 8 % pr. år som er beregnet for torskelerver i Grenland vil bety en nedgang med ca. 55 % i løpet av 10 år. Hvis dette også gjelder i Kristiansandsfjorden, vil det derfor kreve iallfall to, og helst tre års overvåkning å fastslå rimelig sikkert at det har vært en nedgang.
- Det er uansett nødvendig med tre år for å få et inntrykk av hvor stabile nivåene er fra år til år og se om variansanalysen må ta hensyn til tilfeldige avvik fra år til år i gjennomsnitt over områder.
- Det anbefales å ta prøver også fra område E, fordi dette kan komme til å styrke muligheten til å konkludere, men det er ikke tatt i betraktning her i beregningene av utsagnskraft, fordi det er indikasjoner på at tidsutviklingen i område E er forskjellig fra B og D.

**9.2.1.3 Skrubbefilet**

Data fra 1996 finnes for en prøve fra hvert av områdene B og D. Konsentrasjonen i prøven fra område D ligger omtrent 2 lavere enn område B på naturlig logskala, og denne forskjellen er større enn forventet som tilfeldig forskjell ut fra estimert standardavvik (øvre 80 % konfidensgrense 0.5) for enkelprøver fra forskjellige steder og år. Det er derfor antagelig en reell systematisk forskjell mellom områdene, slik at område må være med som faktor i en varians-analyse. For område B finnes også data for en prøve fra 1992. Prøvene fra område B viser en reduksjon på 60 % fra 1996 til 1992, men denne forskjellen er innenfor den statistiske usikkerheten for sammenligning av to enkelprøver, så det er ingen sikker indikasjon på at det har vært en reduksjon.



Hvis ny overvåkning gjennomføres for områdene B og D over 1-3 år, og under samme forutsetninger som beskrevet for forskelever, kan evnen til å påvise endringer anslås til å bli som vist i **Tabell 29**. Hvis en antar at reduksjonsrate i Grenlandsområdet fra 1990 til 2000, som er estimert til å være mindre enn 6 % pr. år, er det lite trolig at tre års overvåkning av område B og D vil være nok til å påvise reduksjon i skrubbefilet fra 1996 til 2006.

**Tabell 29.** Evne til å påvise reduksjon i dioksinkonsentrasjon siden 1996 i skrubbefilet fra Kristiansandsfjorden ved overvåkning av område B og D som funksjon av omfang og lengde på ny overvåkning ved ensidig test for reduksjon med signifikansnivå  $\alpha=0.1$ . Basert på sammenligning av  $n_1=2$  prøver fra periode 1 (år 1996) med  $n_2$  prøver fra periode 2 (Nå); én prøve fra hvert område og år. Det er forutsatt at avvik fra gjennomsnitt for område og periode er uavhengige for enkeltprøver.

Varighet av ny overvåkning:	Ett år	To år	Tre år
Antall nye prøver fra område B+D	1+1	2+2	3+3
$n_2$	2	4	6
$k_F$	2.00	2.67	3.00
$v$	1	3	5
$k_t$	1.00	1.16	1.22
Reduksjon som kan påvises med teststyrke 90 %	87.6 %	65.5 %	60.4 %
Reduksjon som kan påvises med teststyrke 80 %	81.0 %	58.6 %	53.7 %

Analyse av utsagnskraft for test på om nivåer i skrubbefilet er redusert siden 1996 viser:

- For å påvise endring etter ett års overvåkning av skrubbefilet fra område B og D med ensidig test for reduksjon som skissert må konsentrasjonene ha vært redusert med anslagsvis 80-90 % siden 1996.
- Først etter tre års overvåkning kan en regne med på å kunne påvise endring dersom reduksjonen i løpet av 10 år har vært omkring 50-60 %.
- Basert på data fra Grenland kan reduksjonsraten antas å ha vært mindre enn 6 % pr. år, eller mindre enn 45 % på 10 år. Det er derfor tilsomt om en vil kunne påvise noen reduksjon i skrubbeflyndre etter bare 3 års overvåkning.
- Fordi det uansett vil være snakk om svært lite data fra 1996 som sammenligningsgrunnlag, er disse vurderingene spesielt usikre.

9.2.1.4 Krabbesmør

Det finnes data for fire prøver av krabbesmør fra 1996; fra område B, D, E og G. Prøvene fra område D og E har nokså like konsentrasjoner, omtrent 2 lavere på logskala, tilsvarende et forholdstall 6-8 mellom konsentrasjonene, dvs. det samme som for skrubbefilet. Dette er mye mer enn forventet som en tilfeldig forskjell og tyder på det er en reell, systematisk forskjell mellom områdene over tid. Fra område B og D finnes også data fra 1992; det er ingen tegn til nedgang i konsentrasjon fra 1992 til 1996 i krabbesmør fra disse områdene (**Figur 15**) ut fra de få prøvene som foreligger.

Hvis ny prøvetaking gjennomføres fra område B, D og E for sammenligning med data fra 1996 med samme forutsetninger som ovenfor, fås resultater som er vist i **Tabell 30** mht. teststyrke for å påvise endringer.

**Tabell 30.** Evne til å påvise reduksjon i dioksinkonsentrasjon siden 1996 i torskelever i Kristiansandsfjorden ved overvåkning av område B, D og E som funksjon overvåkningsprogrammets lengde ved ensidig test for reduksjon med signifikansnivå  $\alpha=0.1$ . Basert på sammenligning av  $n_1=3$  prøver fra periode 1 (år 1996) med  $n_2$  prøver fra periode 2 (Nå)

Varighet av ny overvåkning:	Ett år	To år	Tre år
Antall nye prøver fra område B,D og E	1+1+1	2+2+2	3+3+3
$n_2$	3	6	9
$k_r$	3.00	4.00	4.50
$v$	2	5	8
$k_i$	1.22	1.41	1.50
Reduksjon som kan påvises med teststyrke 90 %	67.5 %	55.1 %	51.6 %
Reduksjon som kan påvises med teststyrke 80 %	60.4 %	48.6 %	45.3 %

Analyse av utsagnskraft for test på om nivåer i skrubbefilet er redusert siden 1996 viser:

- For at en skal kunne påvise reduksjon fra 1996 noenlunde sikkert med ensidig testing etter ett års overvåkning av krabbesmør i område B, D og E, må konsentrasjonene være redusert med anslagsvis 60-70 % fra 1996 til 2006.
- Etter to års overvåkning kan en regne med å påvise reduksjon på 50-55 % og etter tre år reduksjon på omkring 45-50 %.
- I Grenland ser det ut til at reduksjonraten i perioden etter de store utslippsreduksjonene har vært ca. 9 % pr. år. dvs. 60 % over 10 år. Hvis det samme gjelder i Kristiansandsfjorden etter 1996, er det rimelig å anta at en vil kunne påvise endring fra 1996 etter to års ny overvåkning i Kristiansandsfjorden.
- Det er viktig at fordeling på hann- og hunnkrabbe i prøvene registreres nøyaktig i datamaterialet, helst ved at det analyseres separate blandprøver av begge kjønn. Fett-% bør måles som kovariat, selv om det ikke har så stor betydning som kjønn.

### 9.2.2 Utsagnskraft for tester på om konsentrasjonene fortsatt reduseres

For å teste om konsentrasjonene fortsatt reduseres over tid og anslå hvor raskt det eventuelt skjer er det aktuelt å se bare på data fra en ny overvåkning. Det antas at data samles fra  $j=1, \dots, S$  stasjoner og det tas  $R$  replikate prøver fra hver stasjon innenfor et år. Det antas videre at innsamlingen av materiale til én prøve kan fordelles på  $M$  lokaliteter og/eller tidspunkter på en slik geografisk avstand og tidskala innenfor et år at varians  $\sigma_s^2$  for interaksjon stasjon\*år også gjelder for variasjon i forventningsverdier mellom de  $M$  prøvetakingsstifellene. Det vil bidra til å redusere varians for tilfeldige avvik fra trend for blandprøvene sammenlignet med om alt materialet til en prøve ble samlet på samme sted og tid. Det er usikkert i hvor stor grad dette kan oppnås i praksis i Kristiansandsfjorden, fordi det er så lite data å vurdere det mot, og fordi det kan være avhengig av konkrete forhold innenfor hvert område. Det er likevel tatt med for å vise en prinsipiell mulighet. Det mest realistiske er kanskje å anta  $M=1$ .

Vi har da  $S \cdot R$  analyserte prøver, men  $S \cdot R \cdot M$  innsamlinger av data. Varians av residualer ( $\varepsilon_{ij}$ ) for år  $i$  og stasjon  $j$  sett i forhold til forventningsverdi for hvert år blir

$$Var(\varepsilon_{ij}) = \left( \frac{1}{M} \sigma_{st}^2 + \frac{1}{R} \sigma_r^2 \right) \quad (5)$$

Det teoretiske grunnlaget er skissert mer i detalj i Vedlegg B til og med Vedlegg E. To alternative analysenåter være aktuelle.

- **Alternativ I – Analyse av enkel tidsserie etter aggregering til gjennomsnitt over stasjoner.**

Hvis avvikene fra jevn tidstrend viser stor grad av parallell variasjon på forskjellige stasjoner, bør analysen for tidstrend skje på en enkel tidsserie hvor observasjonene for det enkelte år er aggregert til middelerverdi over stasjoner ( $y_i$ ). Variansen for residualer  $\varepsilon_i$  for årsmidler i forhold til tidstrend estimeres med  $n-2$  frihetsgrader og estimatet har forventningsverdi:

$$Var(\varepsilon_i) = \sigma_i^2 + \frac{1}{S} (Var(\varepsilon_{ij})) \quad (6)$$

Overvåkningsprogrammet antas å bestå av data fra  $n$  år, med  $h$  års mellomrom, dvs. med  $n$  sett av verdier over et tidsrom  $T = (n-1)p + 1$  år, slik at  $x_i = (i-1)h + 1$  for  $i=1, \dots, n$ .

Tidsserien kan skrives:

$$y_i = y_0 + b \cdot (x_i - \bar{x}) + \varepsilon_i; \quad i = 1, \dots, n \quad \text{med} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (7)$$

Utsagnskraften for å teste regresjonskoeffisient  $b$  mot nullhypotese  $b \geq 0$  (dvs. ingen reduksjon) er en funksjon av en såkalt desentralitetsparameter  $\delta$ , som er forholdstallet mellom den faktiske trenden  $b$  og usikkerheten i estimatet for  $b$  (se Vedlegg C):

$$\delta = \frac{-b}{\sqrt{Var(\varepsilon_i)}} h \sqrt{\frac{(n-1)n(n+1)}{12}} \quad (8)$$

- **Alternativ II: Kovariansanalyse med stasjon som faktor - antar uavhengige residualer for hver kombinasjon stasjon\*år**

Antagelsen om uavhengige residualer for hver kombinasjon stasjon\*år ligger innebygd i en vanlig kovariansanalyse med stasjon som kategorisk faktor og år som kovariat. Analysen av data fra Grenland tyder på at dette er ganske godt oppfylt. Datasettet for analyse av tidstrend betraktes som da som  $n \cdot S$  uavhengige observasjoner, og interaksjonen Stasjon\*År brukes som feil-ledd. Det tilsvarer tidtrendanalyse på et datasett med  $S$  uavhengige observasjoner pr. år, hvor årsværdiene for hver stasjon ( $\hat{y}_{st}$ ) er justert for gjennomsnittlig forskjell mellom stasjonene. Variansen for residualene i et slikt datasett blir estimert med  $nS - (S-1) - 2$  frihetsgrader og har forventningsverdi  $Var(\varepsilon_{ij})$  som definert foran. Med samme overvåkningsprogram som beskrevet ovenfor har vi da dataserien:

$$y_{ij} = y_0 + b \cdot (x_i - \bar{x}) + \varepsilon_{ij}; \quad i=1, \dots, n; \quad j=1, \dots, S \quad \text{med} \quad \bar{x} = \frac{1}{nS} \sum_{i=1}^n \sum_{j=1}^S x_{ij} \quad (9)$$

Desentralitetsparameteren som brukes for å estimere utsagnskraft for å teste regresjonskoeffisienten blir nå (se Vedlegg C):

$$\delta = \frac{-b}{\sqrt{\text{Var}(\varepsilon_{ij})}} h \sqrt{\frac{S(n-1)n(n+1)}{12}} \quad (10)$$

Som et tredje alternativ kan det være aktuelt å benytte kovariansanalyse, men justere resultatet for samvariasjon mellom stasjoner. En fremgangsmåte for å gjøre dette er skissert i Milliken og Johnson (1992). Resultatet vil ligge et sted mellom alternativ I og II; alternativ I er omtrent ekvivalent med den mest konservative måten å gjøre denne justeringen på.

Alle resultatene nedenfor er basert på de varianskomponentene som er satt opp i kapittel 9.1. Det vil si at forskjellen mellom alternativene ikke belyser forskjell mellom ulike variansmønstre, men viser forskjellen på å velge en mer eller mindre konservativ analysemetode. For program med flere stasjoner vil teststyrken være følsom for i hvor stor grad avvikene fra tidstrend er uavhengige på forskjellige stasjoner. Ut fra estimerte varianskomponenter ser det ut til at dette kan gjøre seg mest gjeldende for krabbe, og betyr at fordelene av å øke antall stasjoner blir mindre enn beregnet i tabellen.

Ut fra dette er evnen til å påvise endring over tid beregnet for overvåkingsprogram med ulike kombinasjoner av  $n$ ,  $h$ ,  $S$ ,  $M$  og  $R$  for de to alternative analysemåtene. **Tabell 31** viser for en del kombinasjoner av overvåkingsprogram og omfang hvor mye konsentrasjonen må reduseres pr. år for at reduksjonen skal påvises med teststyrke 90 % dersom vi samtidig krever et signifikansnivå på  $\leq 10$  % og bruker en ensidig test hvor nullhypotesen er at konsentrasjonene er konstante eller øker i tid.

Tabellen inneholder også en beregning av kostnader ved ulike kombinasjoner, basert på kostnadskomponenter for prøvetaking og analyse av torsk og krabbe som er vist i **Tabell 32**. Tabellen beskriver også hvordan de totale kostnadene beregnes ved å kombinere kostnadselementene.

Vurderingene av utsagnskraft er gjort for ett sett av varianskomponenter, uten å skille mellom ulike prøvetyper, men når det gjelder kostnader presenteres separate tall for krabbe og torskelever. Som diskutert nedenfor må kostnadene bare ses som regneeksempler, siden de kostnadselementene som inngår vil avhenge av hvordan overvåkingen organiseres helt konkret.

**Tabell 31.** Minste tidstrend (reduksjon pr år) som kan detekteres med teststyrke 90 % og signifikansnivå 10 % for ulik design og varighet av ny overvåking under forutsetninger gitt i teksten. Basert på antatt loglineær tidstrend, dvs. lineær trend på logskala.

*n*: Antall år med observasjoner

S: Antall stasjoner

*h*: Antall år mellom observasjonsår

M: Antall prøvesteder pr. stasjon

*T*: Tidsrom (år) for overvåkings-program,  $(n-1) \cdot h + 1$

R: Antall reprikate prøver pr. stasjon og år

Rad	Omfang av prøvetaking						Reduksjon pr. år som kan detekteres		Regneeksempler for sammenligning av kostnader (1000 kr)			
	n	h	S	M	R	T	alternativ I	alternativ II	Torsk		Krabbe	
									Totalt	pr.år	Totalt	pr.år
1	4	1	4	1	1	4	37.7 %	23.9 %	408	102	316	79
2	5	1	4	1	1	5	23.8 %	16.7 %	510	102	395	79
3	6	1	4	1	1	6	17.1 %	12.5 %	612	102	474	79
4	7	1	4	1	1	7	13.1 %	9.9 %	714	102	553	79
5	8	1	4	1	1	8	10.5 %	8.0 %	816	102	632	79
6	9	1	4	1	1	9	8.7 %	6.7 %	918	102	711	79
7	10	1	4	1	1	10	7.3 %	5.7 %	1020	102	790	79
8	10	1	2	1	1	10	9.4 %	8.2 %	560	57	430	43
9	11	1	3	1	1	11	6.9 %	5.7 %	886	81	688	63
10	13	1	2	1	1	13	6.3 %	5.5 %	728	57	559	43
11	16	1	2	1	1	16	4.5 %	4.0 %	896	57	688	43
12	8	1	8	1	1	8	8.6 %	5.6 %	1552	194	1208	151
13	9	1	8	1	1	9	7.1 %	4.7 %	1746	194	1359	151
14	6	2	4	1	1	11	8.5 %	6.3 %	612	56	474	43
15	7	2	4	1	1	13	6.6 %	4.9 %	714	55	553	43
16	5	3	4	1	1	13	7.9 %	5.6 %	510	39	395	30
17	6	3	4	1	1	16	5.7 %	4.2 %	612	38	474	30
18	3	5	4	1	1	11	18.1 %	7.8 %	306	28	237	22
19	4	5	4	1	1	16	7.5 %	4.8 %	408	26	316	20
20	5	5	4	1	1	21	4.8 %	3.3 %	510	24	395	19
21	10	1	3	1	2	10	7.2 %	5.6 %	1420	142	1120	112
22	10	1	3	2	1	10	7.4 %	5.9 %	940	94	640	64
23	10	1	2	2	2	10	7.3 %	5.8 %	1180	118	810	81

**Tabell 32.** Eksempel på mulige kostnadskomponenter ved overvåkning av dioksin i blåskjell, torsk og krabbe. Basert på opplysninger for JAMP-programmet.

$n$ : Antall år med observasjoner

$M$ : Antall prøvesteder pr. stasjon

$I$ : Antall individer pr. prøve

$S$ : Antall stasjoner

$R$ : Antall replikate prøver pr. stasjon og år

Kostnadstype:	Kostnad (1000 kr)				Faktor i totalkostnad
	Blåskjell	Torsk	Krabbe		
Eksterne kostnader pr. tokt	$KE_{Ttokt}$	5	5	2	$n$
Adm.kostnader pr. tokt:	$KA_{Ttokt}$	5	5	5	$n$
Adm.kostnader pr. stasjon	$KA_{Stasjon}$	2	2	1	$n \cdot S$
Eksterne kostnader pr. prøvested og tokt	$KE_p$	5	5	1	$n \cdot S \cdot M$
Prøveopparbeiding	$K_{Prøve}$	1	1	1	$n \cdot S \cdot R$
Prøveopparbeiding pr. individ i prøven	$K_{Individ}$	0.25	0.25	0.25	$n \cdot S \cdot R \cdot I$
Kjemisk analyse	$K_{analyse}$	10	10	10	$n \cdot S \cdot R$

Total kostnad =  $n \cdot (KE_{Ttokt} + KA_{Ttokt} + S \cdot (KA_{Stasjon} + M \cdot KE_p + R \cdot (K_{Prøve} + I \cdot K_{Individ} + K_{analyse}))$

Kostnadselementene kan ikke angis nøyaktig som generell gyldige; de må bare ses som realistiske eksempler på mulige kostnader. Spesielt tokt- og stasjonskostnader vil kunne variere sterkt fra gang til gang avhengig av mange faktorer, for eksempel om prøver samles inn lokalt eller ved tokt av forskningspersonell, om det er hovedmål med toktet eller en deloppgave. Båt- og mannskaps-kostnader, samt geografiske forhold som stasjonsavstand og evt. reisekostnader vil også kunne variere.

Kostnader til prøveopparbeiding er basert på erfaringstall for tidsforbruk og NIVAs timepriser, men vil sels sagt også kunne variere. For blåskjell er det regnet med et tidsforbruk på 3 timer for en prøve på 60 skjell; tidsforbruket kan variere fra ca. 2 til 4 timer, bl.a. avhengig av om prøvene skal måles og sorteres etter størrelse. For torsk er det regnet med et tidsforbruk til prøveopparbeiding på ca. 8 timer for 20 torsk, og tilsvarende for krabbe. Det er antatt en timepris på 750 kr./time.

Hvis vi går ut fra at en statistisk analyse for å påvise tidstrend vil bli basert på alternativ I ovenfor, med analyse av tidsserie aggregert til gjennomsnitt over stasjoner, og holder antall stasjoner på 4, som kan være realistisk å oppnå, kreves overvåkning hvert år i 8-10 år eller mer før en kan regne med å påvise noenlunde sikkert tidstrender omtrent som de som forekommer i tidligere data ved en så konservativ metode<sup>1</sup> (rad 1-7). Hvis testene av samme data kan baseres på alternativ II, dvs. hvor data kommer fra 4 stasjoner med uavhengige avvik fra trenden mellom alle kombinasjoner av stasjon og år, vil en tidstrend av samme størrelse kunne påvises ett til to år før.

I de etterfølgende radene i tabellen er ulike dimensjoner i et overvåkningsprogram justert på forskjellige måter for å gi alternativer med omtrent sammenlignbare krav til tidstrend, slik at de viser hvordan det innvirker på krav til varighet for å påvise samme faktiske trend, og med regneeksempler for hvilke kostnader det medfører.

<sup>1</sup> I de kovariansanalysene som er gjort i denne rapporten er det ikke brukt en så konservativ metode, i stedet er det antatt at residualene for hver kombinasjon av stasjon og år er uavhengige av hverandre.

Hvis antall stasjoner reduseres fra 4 til 2 vil det ta 2-3 år mer å oppdage samme trend. Virkningen blir størst for alternativ II med analyse av enkelldata. Hvis antall stasjoner reduseres tar det lengre tid å oppdage en reduksjon. Som regneeksemplene i tabellen viser, kan det til gjengjeld være realistisk å anta at totalkostnadene blir hhv. 15 og 30 % lavere enn med 4 stasjoner.

Hvis antall stasjoner dobles fra 4 til 8, vil samme trend kunne detekteres ca. ett år tidligere for alternativ I, og ca. to år før for alternativ II, men samtidig øker totalkostnad ifølge regneeksemplene i tabellen med hhv. 70 og 50 % (rad 11 og 12 sammenlignet med rad 7). Det er usikkert hva en vinner med så mange stasjoner over et begrenset område. Jo flere stasjoner, jo lettere vil de ligge, det kan da være at de vil få større samvariasjon. Det er ikke sikkert det er realistisk med mer enn 3-4 stasjoner med uavhengige residualer.

Resultatene i rad 1-12 er basert på årlig gjennomføring av overvåkingsprogrammet. Hvis det i stedet gjennomføres overvåking bare hvert andre, tredje eller femte år, vil det kreves lengre tid å påvise en trend, men kostnadene blir lavere med de kostnadselementene som er brukt i eksemplene. Med prøvetaking hvert annet år kreves et ca. 2 år lenger tidsrom for å detektere en reduksjonsrate på ca. 6 % årlig (rad 6 mot 13). Med prøvetaking hvert tredje år vil tidsrommet bli ca. 3 år lenger. Hvis prøvetaking gjøres bare hvert femte år vil det kunne gå 16 til 21 år før en detekterer en trend som med årlig prøvetaking påvises etter 10 år (rad 17 og 18). Alle disse resultatene gjelder med 4 stasjoner. Det tar altså lengre tid, men i følge regneeksemplene vil totale kostnader for et overvåkingsprogram som kan detektere tidstrend på ca. 7 % årlig vil reduseres med hhv. 30, 50 og 60 % ved prøvetaking hvert annet, tredje eller femte år, sammenlignet med årlig prøvetaking.

De siste radene i tabellen viser ulike kombinasjoner av S, M og R som gir omtrent samme evne til å detektere trender i løpet av 10 års årlig prøvetaking som et enkelt program med 4 stasjoner. Ved å analysere to replikater fra hver stasjon ( $R=2$ ) eller basere hver blandprøve på to prøver fra ulikt sted eller tid innen år og stasjon som beskrevet tidligere ( $M=2$ ) kan en oppnå omtrent samme deteksjons-evne med tre stasjoner, og gjøres begge deler kan det oppnås med to stasjoner. På grunn av høye analysekostnader er bare alternativet med  $S=3$ ,  $M=2$  og  $R=1$  mer kostnadseffektivt enn å ta enkeltprøver fra 4 stasjoner, og forskjellen er 5-10 % av totale kostnader. Siden kostnadsanslagene er noe usikre, gir det ingen tydelig forbedring.

Teststyrken er en funksjon av hvor store endringene er. Hvis endringene er mindre enn de som er angitt i tabellene, vil de påvises med teststyrker under 90 %. F.eks. vil endringer på 83 % av de som er vist i tabellen påvises med teststyrke 80 % i stedet for 90 % for ensidige tester med  $\alpha=0.1$ .

Hvis en i tillegg til å påvise at det er en ending (reduksjon) også vil ha et brukbart estimat for hvor raskt det går, kan kravet til overvåkingsperiode. Nøyaktigheten i estimatet for tidstrend, uavhengig av hvor stor trenden er, vil selvsagt stå i forhold til evnen til å påvise at det skjer en reduksjon. Grovt sagt vil et 90 % konfidensintervall for estimert tidstrend ha en bredde som tilsvarer omtrent den nedre grensen for påvisbar trend i **Tabell 31**. For eksempel vil årlig overvåking på 4 stasjoner i 10 år kunne påvise en felles trend på 7.3 % hvis data aggregeres til en enkel tidsserie (rad 7, alternativ D). Da vil 90 % konfidensintervall være estimert trend  $\pm 3.7$  %. Hvis en vil ha et mer nøyaktig estimat, må perioden være lengre. Hvis den forlenges til 15 år kan en regne med at 90 % konfidensintervall vil bli  $\pm 2.0$  % pr. år rundt estimert trend.

Hvis testene gjøres mer konservative ved at en bruker  $\alpha=0.05$  for ensidige tester, vil endringene måtte være 15 til 40 % større enn vist i **Tabell 31** for at de skal påvises med teststyrke 90 % for de forskjellige prøveprogrammene.

Analyse av utsagnskraft for tester på om konsentrasjonene fortsatt reduseres gir følgende resultat:

- En tidstrendanalyse på årsmidler over 4 stasjoner vil kreve overvåkning i 8-10 år eller mer før en kan regne med å påvise noenlunde sikkert en reduksjonstrate på 5-10 % pr. år. En slik reduksjonstrate tilsvare omtent det en finner i tidligere data. Hvis testene kan baseres på at avvikene for hver kombinasjon av stasjon og år i stor grad er uavhengige av hverandre, vil en tidstrend av samme størrelse kunne påvises ett til to år før.
- Med 3 stasjoner i stedet for 4 vil det ta ett år mer å oppdage samme trend, og med bare 2 stasjoner vil det ta 2-3 år lenger. Samtidig kan totalkostnadene frem til en kan påvise reduksjonen bli hhv. 15 og 30 % lavere. En økning fra 4 til 8 stasjoner kan gi ett til 2 års raskere påvisning av en trend, men samtidig øker total kostnad med hhv. 70 og 50 %. Det er usikkert hva en vinner med så mange stasjoner over et begrenset område, siden de kan ha større samvariasjon jo tettere de ligger. For Kristiansandsfjorden kan 3-4 stasjoner være et optimalt antall med hensyn til en slik vurdering.
- Med prøvetaking hvert annet år kreves et 2 år lenger tidsrom før en kan påvise en reduksjon på ca. 6 % pr. år like sikkert som med årlig prøvetaking; med prøvetaking hvert tredje år kreves et 3 år lenger tidsrom. Hvis det tas prøver bare hvert femte år kan det ta 5 til 10 år mer før en kan påvise samme trend som for årlig overvåkning etter ca. 10 år. Alle disse sammenligningene gjelder 4 stasjoner.
- Regneeksempler på kostnader ved overvåkning av konsentrasjoner i torsk og krabbe anwyder 30-60 % mindre totale kostnader for alternativene med flere år mellom hver prøvetaking sammenlignet med årlig prøvetaking i forhold til den utsagnskraften som oppnås, til gjengjeld vil det gå lenger tid før en oppnår samme utsagnskraft for å påvise endring. Optimalisering av overvåkning mhp. kostnader må vurderes nærmere i sammenheng med konkret planlegging av overvåkning, når ulike kostnadselementer er fastlagt.
- Dersom en kan finne 6 prøvetakingssteder med uavhengige variasjoner, og analyserer 3 blandprøver som hver er basert materiale fra to steder, kan det gi samme utsagnskraft som enkelprøver fra 4 stasjoner til ca. 10 % mindre kostnad over samme tidsrom, men denne forskjellen må betraktes som usikkert anslått.

### 9.2.3 Utsagnskraft for å teste om gjennomsnittskonsentrasjoner ligger innenfor gitte klassifiseringsgrenser.

Det kan også være aktuelt å bruke nye overvåkingsdata til å klassifisere et geografisk område ved å sammenligne estimat for gjennomsnittskonsentrasjon  $\mu$  med en klassifiseringsgrense, for eksempel for å revurdere kostholdsråd.

Nullhypotesen  $H_0$  er at gjennomsnittlig konsentrasjon  $\mu$  overskrider en klassifiseringsgrense  $\mu_C$ . Hvis grensen ikke er overskredet, men det faktisk gjennomsnittet er  $\mu=f\mu_C$  hvor  $f < 1$ , ønsker vi at  $H_0$  skal forkastes og området klassifiseres som å være innenfor grensen med så stor sannsynlighet som mulig. Transformert til logaritmisk skala<sup>1</sup> tilsvare det at  $\log(\mu) = \log(f) + \log(\mu_C)$ . Spørsmålet er hvor langt

<sup>1</sup> Klassifiseringsgrensene vil være definert som gjennomsnitt på lineær skala, men for enkelhets skyld ser vi her bort fra det. For å regne om en klassifiseringsgrense på lineær skala til grenseverdi for gjennomsnitt på logskala må det legges inn en faktor som kompenserer for skjevfordelingen. Det innebærer at grensen på logskala blir noe mindre enn logaritmen av grenseverdien på lineær skala.



under grensen den faktiske konsentrasjonen må ligge for at nullhypotesen skal bli forkastet med gitt teststyrke 1- $\beta$  i en ensidig test med definert signifikansnivå for et visst overvåkingsprogram.

Dette fungerer som en såkalt *grønn* test, dvs. den er konservativ når det gjelder å klassifisere tilstanden til å oppfylle et gitt kvalitetskriterium, og stor usikkerhet pga. lite data vil gi mindre sjanse for å oppnå dette for en gitt faktisk tilstand. Metoden vil gi liten sannsynlighet (=signifikansnivået) for å godkjenne vannforekomster som ligger like under grensen. Hvis dette anses for strengt, kan det eventuelt justeres ved å legge inn en viss margin eller toleranse i forhold til formell klassifiseringsgrense<sup>1</sup>, men det berører ikke vurderingen av utsagnskraft her.

Hvis det er lite reduksjon pr. år, er det rimelig å klassifisere ut fra forskjellen mellom klassifiseringsgrensen og et gjennomsnitt over alle data. Hvis overvåkingen har pågått over et antall år, og det er en tydelig nedgang i løpet av perioden, kan det i stedet være aktuelt å klassifisere etter estimert forventningsverdi på regresjonskurven ved slutten av overvåkingsperioden. I begge tilfellene blir usikkerheten i estimatene tatt i betraktning. Det velges et signifikansnivå som en akseptabel risiko for å konkludere med at gjennomsnittet er innenfor klassifiseringsgrensen hvis det i virkeligheten ligger over, og utsagnskraften uttrykkes da som den øvre grense for faktoren *f* som gir ønsket teststyrke for å konkludere med at konsentrasjonene er innenfor klassifiseringsgrensen. Jo større *f* er, jo bedre er testen i stand til å klassifisere, dvs. skille skarpt mellom tilstand over og under grensen. Krav til overvåkingsprogram må da vurderes ut fra hvor lav *f* en kan akseptere for å oppnå gitt teststyrke. Detaljer er beskrevet i Vedlegg E.

Tabell 33 viser estimater for hvor lav den reelle gjennomsnittskonsentrasjonen må være i forhold til klassifiseringsgrensen for at en med 90 % sannsynlighet skal konkludere med at området er innenfor grensen når det brukes en test med signifikansnivå 10 % for å forkaste nullhypotesen. Det er her antatt at både konsentrasjoner og klassifiseringsgrense er logtransformert før analyse, dvs. at klassifiseringsgrensen gjelder for geometrisk middel (se side 26). I tabellen er resultatet for nødvendig forskjell på logskala regnet tilbake til forholdstall mellom geometrisk middel og klassifiseringsgrense for utransformerte konsentrasjoner. Tabellen har med de samme alternativene for design og omfang av overvåking som Tabell 31.

Tabellen indikerer at hvis data aggregeres til en enkel tidsserie før tidstrendanalyse, og det er en stasjonær tilstand hvor konsentrasjonene fluktuerer rundt konstant gjennomsnitt over tid, må reell gjennomsnittlig konsentrasjon være innenfor 0.65 ganger klassifiseringsgrensen hvis 4 års overvåking på 4 stasjoner skal gi spesifisert sikkerhet i klassifiseringen (rad 1 i tabellen). Hvis konsentrasjonene synker over tid, må gjennomsnittet komme ned i 0.5 ganger klassifiseringsgrensen ved slutten av perioden for å oppnå dette etter 4 års overvåking.

Med 10 års overvåking hvert år på 4 stasjoner oppnås samme klassifiseringsikkerhet hvis konsentrasjoner er mindre enn 0.8 ganger klassifiseringsgrensen uten tidstrend, eller synker til under 0.7 ganger denne grensen i løpet av overvåkingen (rad 7 i tabellen). Med data fra bare to stasjoner kreves 16 års overvåking for å oppnå samme evne til å klassifisere, selv om forskjellen på 10 og 16 års overvåking er relativt liten (rad 8-10 sammenlignet med rad 7).

---

<sup>1</sup> Spørsmålet om justering av grenseverdi må åpenbart ses i sammenheng med hva grenseverdien er tenkt å gjelde (middelverdi, ulike persentiler) og hvordan en anser at konsentrasjonene er fordelt, for eksempel om de er log-normalt fordelt.

**Tabell 33.** Største forholdstall  $f$  mellom virkelig gjennomsnittskonsentrasjon og klassifiseringsgrense for konsentrasjon  $\mu_C$  som gir konklusjon 'gjennomsnitt  $< \mu_C$ ' med 90 % sannsynlighet for ensidig test med signifikansnivå 10 % for ulik design og varighet av ny overvåking. Kombinasjonene er de samme som i **Tabell 31**.

Omfang av prøvetaking							Forholdstall $f =$ konsentrasjon/klassifiseringsgrense			
							Alternativ I: enkel tidsserie: Gjennomsnitt over stasjoner		Alternativ II: kovarians-analyse med stasjon som faktor	
							Test på gjennomsnitt ( $\bar{y}$ ) over alle data	Test på estimert verdi ( $\hat{y}_n$ ) for siste tidspunkt	Test på gjennomsnitt ( $\bar{y}$ ) over alle data	Test på estimert verdi ( $\hat{y}_n$ ) for siste tidspunkt
Rad	n	h	S	M	R	T				
1	4	1	4	1	1	4	0.66	0.49	0.77	0.72
2	5	1	4	1	1	5	0.71	0.56	0.79	0.75
3	6	1	4	1	1	6	0.75	0.60	0.81	0.77
4	7	1	4	1	1	7	0.77	0.62	0.82	0.78
5	8	1	4	1	1	8	0.79	0.64	0.83	0.79
6	9	1	4	1	1	9	0.80	0.66	0.84	0.80
7	10	1	4	1	1	10	0.81	0.68	0.85	0.81
8	10	1	2	1	1	10	0.76	0.60	0.79	0.70
9	11	1	4	1	1	11	0.80	0.66	0.82	0.76
9	13	1	2	1	1	13	0.79	0.64	0.81	0.73
10	16	1	2	1	1	16	0.81	0.67	0.83	0.76
11	8	1	8	1	1	8	0.82	0.70	0.88	0.86
12	9	1	8	1	1	9	0.83	0.71	0.89	0.87
13	6	2	4	1	1	11	0.75	0.60	0.81	0.77
14	7	2	4	1	1	13	0.77	0.62	0.82	0.78
15	5	3	4	1	1	13	0.71	0.56	0.79	0.75
16	6	3	4	1	1	16	0.75	0.60	0.81	0.77
17	3	5	4	1	1	11	0.48	0.31	0.73	0.69
18	4	5	4	1	1	16	0.66	0.49	0.77	0.72
19	5	5	4	1	1	21	0.71	0.56	0.79	0.75
20	10	1	3	1	2	10	0.81	0.68	0.85	0.80
21	10	1	3	2	1	10	0.81	0.67	0.84	0.80
22	10	1	2	2	2	10	0.81	0.68	0.84	0.78

Hvis en i stedet kan analysere data for enkelprøver i en kovariansanalyse med stasjon som faktor, dvs. hvis stasjonene har uavhengige fluktusjoner over tid i forhold til trenden, oppnås samme sikkerhet i klassifiseringen ved konsentrasjoner som kan være noe høyere opp mot grensen enn for analyse av en enkel tidsserie av års-gjennomsnitt over stasjoner. Forskjellen er størst dersom en tester estimert sluttkonsentrasjon mot klassifiseringsgrensen, og for en relativt kort overvåkingsperiode. Med 4 stasjoner som har uavhengige fluktusjoner oppnås stort sett ønsket klassifiseringsikkerhet dersom konsentrasjonene er innenfor 0.7 til 0.85 ganger klassifiseringsgrensen.

Hvis en ikke tar prøver hvert år, vil klassifiseringssikkerheten bli dårligere for overvåking over samme totale tidsrom dersom det velges å aggregere til én tidsserie før tidsserieanalyse. Hvis en for eksempel tar prøver hvert 5. år på 4 stasjoner og konsentrasjonene synker over tid, må det reelle

gjennomsnittet komme ned under en tredjedel av klassifiseringsgrensen for å oppnå ønsket klassifiseringssikkerhet etter et tidsrom på 11 år (rad 17). Hvis data kan analyseres som prøver med uavhengige fluktasjoner på hver stasjon i forhold til gjennomsnitt eller trend (Alternativ II), har dette mindre å si enn en tidsserie av gjennomsnitt over stasjoner.

Klassifisering i forhold til grenseverdier på konsentrasjon bør gjøres ved ensidig test hvor en konkluderer med at klassifiseringskravet er oppfylt bare dersom gjennomsnittskonsentrasjon ligger klart under klassifiseringsgrensen, dvs. at må reell gjennomsnittskonsentrasjon må være noe mindre enn grenseverdien. Med signifikansnivå 10 % (risiko for å godta at klassifiseringskriteriet er oppfylt hvis det ikke er tilfelle), gjelder følgende:

- Med en enkel serie av årsmidler basert på data over 4 år fra 4 stasjoner må forholdstallet mellom reell gjennomsnittskonsentrasjon og klassifiseringsgrensen være mindre enn 0.66 for at teststyrken skal bli 90 % hvis målte konsentrasjoner varierer tilfeldig rundt et konstant gjennomsnitt uten noen systematisk ending over tid. Teststyrken er sannsynligheten for å konkludere riktig dersom konsentrasjonen virkelig er under klassifiseringsgrensen.
- Hvis konsentrasjonene reduseres systematisk over tid, bør klassifiseringen gjøres ved å sammenligne grenseverdien med estimert gjennomsnitt ved slutten av 4-årsperioden. Med en serie av årsmidler basert på 4 stasjoner må forholdstallet da komme ned i under 0.49 ved slutten av perioden for å oppnå samme teststyrke.
- Etter 10 års overvåkning hvert år på 4 stasjoner oppnås samme teststyrke for forholdstall som er noe høyere, opp mot 0.8 uten noen tidstrend og 0.65 ved slutten av perioden hvis det er en tidstrend. Med data over 10 år fra bare to stasjoner kreves 6 års lenger overvåkning for å oppnå samme evne til å klassifisere, selv om forskjellen mellom 10 og 16 års overvåkning ikke er så stor.
- Hvis stasjonene har stor grad av uavhengige avvik fra trend, slik at de kan analyseres som uavhengige observasjoner fra ulike kombinasjoner av stasjon og år, kan evnen til å klassifisere riktig forbedres. Det gjelder spesielt for de kombinasjonene som krever lavt forholdstall når en tester på en enkel tidsserie aggregert over stasjoner. Med data over 4 år fra 4 stasjoner må forholdstallet konsentrasjon:klassifiseringsgrense være under hhv. 0.77 og 0.72 uten og med tidstrend i stedet for 0.66 og 0.49 som gjelder enkel tidsserie som gjennomsnitt over stasjoner. For 10 års overvåkning må forholdstallet være under hhv. 0.85 og 0.81 i stedet for 0.81 og 0.65. Dette betyr mest for det tilfelle at det er en signifikant tidstrend, slik at det vil være naturlig å klassifisere ut fra tilstanden mot slutten av tilgjengelig dataserie i stedet for et generelt gjennomsnitt over hele overvåkningsperioden.

## 10 Litteratur

- Bakke, T.; Schlabbach, M., NILU; Ruus, A.; Skare, J. U. (NVH/VD); Bjerkeng, B.; Berg, V. (NVH/VD); Knudsen, J. A.(HD) (2003): Overvåking av miljøgifter i fisk og skalldyr fra Grenlandsfjordene 2001-2002. NIVA-rapport nr. 4702. ISBN:82-577-4370-4
- Berge, J.A.; Knutzen, J. (1991): Sedimentenes betydning for forurensningstilstanden i Frierfjorden og tilgrensende områder. Rapport 3: Eksperimentelt opptak av persistente klororganiske forbindelser og kvikksølv i skrubbe og krabbe, opptak/utskillelse i blåskjell og registrering av miljøgiftinnhold i bunndyr fra Frierfjorden og Breivikfjorden. NIVA-rapport nr. 2573. ISBN:82-577-1897-1
- Berglind, L.; Brevik, E.M.; Green, N.; Kringstad, A. (SD); Oehme, M. (NILU); Skåre, J.U. (NVH) (1993): Overvåking av miljøgifter i fisk og skalldyr fra Grenlandsfjordene 1991. NIVA-rapport nr. 2833. ISBN:82-577-2231-6
- Bjerkeng, B.; Green, N. (2000): The Voluntary international contaminant-monitoring (VIC) for temporal trends with the aim to test sampling strategies for a co-operative revision of guidelines by 1999. Presentert på møte i OSPAR "Working Group On Concentrations, Trends And Effects Of Substances In the Marine Environment" (SIME), Stockholm 21-25 february 2000.
- Bjerkeng, B.; Ruus A. (2002): Statistisk analyse av dioksiner i organismer i Frierfjord/Grenlandsområdet. TA-nr. 1916/2002. NIVA-rapport 4595-2002. ISBN: 82-577-4255-4.
- Green, N.; Helland, A.; Hylland, K.; Knutzen, J.; Walday, M. (2001): Joint Assessment and Monitoring Programme (JAMP). Overvåking av miljøgifter i marine sedimenter og organismer 1981-1999. NIVA-rapport nr. 4358. ISBN:82-577-3995-2
- Knutzen, J.; Enger, B.; Martinsen, K. (1986a): Basisundersøkelse av Kristiansandsfjorden. Delrapport 4. Miljøgifter i fisk og andre organismer 1982-1984. NIVA-rapport nr. 1848. ISBN:82-577-1056-3
- Knutzen, J.; Martinsen, K. (SD) (1986b): Tiltaksorientert overvåking av miljøgifter i fisk og andre organismer fra Kristiansands-fjorden 1985. NIVA-rapport nr. 1940. ISBN:82-577-1168-3
- Knutzen, J.; Martinsen, K. (SD); Oehme, M. (NILU) (1988a): Tiltaksorientert overvåking av miljøgifter i organismer og sedimenter fra Kristiansandsfjorden 1986-1987. NIVA-rapport nr. 2114. ISBN:82-577-1329-9
- Knutzen, J.; Oehme, M. (NILU) (1988b): Undersøkelse av klorete dioksiner og dibenzofuraner i fisk, skalldyr og sedimenter fra Frierfjorden og tilgrensede områder 1987 - 1988. NIVA-rapport nr. 2189. ISBN:82-577-1477-1
- Knutzen, J. (1990): Polyklorete dibenzofuraner/dioksiner og andre persistente klororganiske forbindelser i sjø-ørret fra Klosterfoss/Skienselva oktober 1989. NIVA-rapport nr. 2393. ISBN:82-577-1684-7
- Knutzen, J.; Oehme, M. (NILU) (1990): Klorete dibenzofuraner og dioksiner i krabber, fisk og reker fra Frierfjorden, tilstøtende områder og referansestasjoner 1988-1989. NIVA-rapport nr. 2346. ISBN:82-577-1629-4
- Knutzen, J.; Green, N. (1991a): Overvåking av miljøgifter i fisk og blåskjell fra Grenlandsfjordene 1990. NIVA-rapport nr. 2636. ISBN:82-577-1963-3
- Knutzen, J.; Martinsen, K. (SD); Næs, K.; Oug, E.; Oehme, M (1991b): Tiltaksorientert overvåking av miljøgifter i organismer og sedimenter fra Kristiansandsfjorden 1988 og 1990. NIVA-rapport nr. 2554. ISBN:82-577-1873-4

- Knutzen, J.; Becher, G. (Folkehelsen); Kringstad, A. (SINTEF/SD); Oehme, M. (NILU) (1994a): Overvåking av miljøgifter i organismer fra Kristiansand 1992. NIVA-rapport nr. 2996. ISBN:82-577-2430-0
- Knutzen, J.; Berglind, L.; Brevik, E.M.; Green, N.; Oehme, M. (NILU); Schlabbach, M. (NILU); Skåre, J.U. (NVH/VI) (1994b): Overvåking av miljøgifter i fisk og skalldyr fra Grenlandsfjordene 1992. NIVA-rapport nr. 2989. ISBN:82-577-2427-0
- Knutzen, J.; Becher, G.; (Folkehelsen); Biseth, A. (NILU); Brevik, E.M.; Green, N.; Schlabbach, M. (NILU); Skåre, J.U. (NVH/VI) (1995a): Overvåking av miljøgifter i fisk og skalldyr fra Grenlandsfjorden 1993. NIVA-rapport nr. 3195. ISBN:82-577-2675-3
- Knutzen, J.; Brevik, E.M.; Biseth, A.; Schlabbach, M. (1995b): Innledende forsøk med utskillelse av polyklorete dibenzofuraner/dibenzo-p-dioxiner, heksaklorbenzen, oktaoklorstyren og deklorklorbifenyli hos forsk (Gadus morhua) og skrubbe (Platichthys flesus) fra Frierfjorden. NIVA-rapport nr. 3328. ISBN:82-577-2855-1
- Knutzen, J.; Brevik, E.M.; Green, N.; Biseth, A. (NILU); Schlabbach, M. (NILU); Skåre, U.J. (NVH/VI) (1995c): Overvåking av Grenlandsfjordene 1994. NIVA-rapport nr. 3363. ISBN:82-577-2860-8
- Knutzen, J.; Biseth, A.; Green, N.; Schlabbach, M.; Skåre, J.U. (1996): Overvåking av miljøgift i fisk og skalldyr fra Grenlandsfjordene 1995. NIVA-rapport nr. 3590. ISBN:82-577-3143-9
- Knutzen, J.; Brevik, E.M.; Green, N.; Biseth, A.; Egaas, E.; Schlabbach, M.; Skåre, J.U. (1998a): Overvåking av miljøgifter i fisk og skalldyr fra Grenlandsfjordene 1996. NIVA-rapport nr. 3834. ISBN:82-577-3414-4
- Knutzen, J.; Næs, K.; Berglind, L.; Brevik, E.M.; Følsvik, N.; Biseth, A. (NILU); Schlabbach, M. (NILU) (1998b): Overvåking av miljøgifter i sediment og organismer fra Kristiansandsfjorden 1996. NIVA-rapport nr. 3833. ISBN:82-577-3413-6
- Knutzen, J.; Becher, G.; Biseth, A.; Bjerkeng, B.; Brevik, E.M.; Green, N.; Schlabbach, M.; Skåre, J.U. (1999a): Overvåking av miljøgifter i fisk og skalldyr fra Grenlandsfjordene 1997. NIVA-rapport nr. 4065. ISBN:82-577-3670-8
- Knutzen, J. Becher, G. Berglind, L. Brettum, P. Schlabbach, M.; Skåre, J.U. (1999b): Organiske miljøgifter i taskekrabbe (Canser pagurus) fra norske referanselokalieter 1996. Undersøkelse av polyklorete dibenzofuraner/dibenzo-p-dioxiner (PCDF/PCDD), andre persistente klororganiske stoffer og polysykliske aromatiske hydrokarboner (PAH). NIVA-rapport nr. 4068. ISBN:82-577-3673-2
- Knutzen, J.; Biseth, A.; Brevik, E.M.; Green, N.; Schlabbach, M.; Skåre, J.U. (2000a): Overvåking av miljøgifter i fisk og skalldyr fra Grenlandsfjordene 1998. NIVA-rapport nr. 4207. ISBN:82-577-3826-3
- Knutzen, J.; Brevik, E.M.; Green, N.; Schlabbach, M.; Skåre, J.U. (2000b): Overvåking av miljøgifter i fisk og skalldyr fra Grenlandsfjordene 1999. NIVA-rapport nr. 4317. ISBN:82-577-3949-9
- Knutzen, J.; Bjerkeng, B. Green, N. Kringstad, A. Schlabbach, M. Skåre, J.U. (2001): Overvåking av miljøgifter i fisk og skalldyr fra Grenlandsfjordene 2000. NIVA-rapport nr. 4452. ISBN:82-577-4098-5
- Milliken & Johnson (1992): Analysis of Messy Data Volume I: Designed Experiments. Chapman & Hall.

## Vedlegg A. Logtransformasjon

De statistiske analysene vil være basert på naturlige logaritmer, både av miljøgift-konsentrasjoner og av biologiske karakteristikk som lengde, vekt og fett-%.

Naturlig logaritme  $\ln(x)$  av et tall  $x$  er definert ved at  $x = e^{\ln(x)}$  hvor konstanten  $e = 2.7182$ . De naturlige logaritmene har den egenskapen at små forskjeller i logaritmene til to verdier gir direkte uttrykk for relativ forskjell, dvs. at  $\ln(x_1) - \ln(x_2) \cong (x_1 - x_2)/x_2$ . For større forskjeller vil sammenheng mellom forskjell av logtransformerte verdier og relative forskjell være som i tabellen nedenfor:

$\ln(x_1) - \ln(x_2)$	$x_1/x_2$
0.1	1.11
0.2	1.22
0.5	1.65
1	2.72
1.5	4.48
2.3	9.97

Logtransformering har vist seg å gi mer homogen varians i fordelingen av residualene ved tilpasning av statistiske modeller, og gjør at faktorer som virker multiplikativt kan analyseres i additive modeller hvor konsentrasjon beskrives som en sum av ulike effekter.

Det virker rimelig at opptak og akkumulering av miljøgifter er en multiplikativ prosess, dvs. at miljøgiftkonsentrasjoner i organismer er et resultat av konsentrasjoner i ytre miljø multiplisert med en konsentrasjonsfaktor som uttrykk for opptaksprosesser og utskilling/nedbryting. Det er naturlig å anta at variasjoner over tid i ytre miljø også er proporsjonalt med det generelle gjennomsnittsnivået i omgivelsene, altså som et produkt av et generelt gjennomsnitt og en faktor som varierer over tid. På samme måte er det mest naturlig å anta at variasjoner som skyldes biologiske faktorer i organismene, f.eks. vekst, gyting, akkumulering og forbrenning av fett etc., i hovedsak gir de samme relative virkninger på miljøgiftkonsentrasjonene, uavhengig av hvor høy konsentrasjonen er. Dette vil ikke være tilfelle hvis nedbrytings og utskillingsmekanismer utløses eller forsterkes ved bestemte konsentrasjonsnivåer eller hvis konsentrasjonene blir så høye at de virker tilbake på de biologiske prosessene. I slike tilfeller vil det være avvik fra proporsjonale virkninger, men det må likevel være en eller annen for multiplikativ samvirke.

## Vedlegg B. Test-scenario – kovariansmodell

Anta at vi skal overvåke konsentrasjonen (C) i organismer i et område ved å ta prøver på flere punkter og gjentatt over tid. Den variasjonen vi vil se i materialet kan oppsummeres i en kovariansmodell:

$$\log(C_{ijk}) = \mu + b \cdot x_i + \varepsilon_i + \alpha_j + \tau_{ij} + \sum_{v=1}^{N_v} \gamma_v z_{v,k(ij)} + \mathcal{E}_{k(ij)} \quad (11)$$

hvor indeks  $i$  angir tidspunkt,  $j$  er sted, og  $k$  replikat prøve innenfor kombinasjon av sted og tidspunkt. Modellen inneholder både en tidstrend, her som en lineær trend på logskala ( $b \cdot x_i$ ), variasjon mellom steder ( $\alpha_j$ ), fluktasjoner over tid i gjennomsnitt over stasjoner ( $\varepsilon_i$ ) og uavhengig for hver stasjon ( $\tau_{ij}$ ), og en sum av korreksjonsledd for  $N_v$  kovariater  $z_v$  som f.eks. kjønn, lengde, vekt eller fett-%. I tillegg er det et feilledd for hver enkelt prøve ( $\mathcal{E}_{k(ij)}$ ).

Vi antar at verdiene korrigeres best mulig for kovariatene, slik at vi får en modell med verdier justert for kovariatene:

$$\begin{aligned} \log(C_{ijk}^*) &= \log(C_{ijk}) - \sum_{v=1}^{N_v} \gamma_v^* (z_{v,k(ij)} - z_{v,0}) \\ &= \mu + b \cdot x_i + \varepsilon_i + \alpha_j + \tau_{ij} + \sum_{v=1}^{N_v} (\gamma_v - \gamma_v^*) (z_{v,k(ij)} + \gamma_v^* z_{v,0}) + \mathcal{E}_{k(ij)} \\ &= \underbrace{\mu + \gamma_v^* x_{v,0}}_{\mu^*} + b \cdot x_i + \varepsilon_i + \alpha_j + \tau_{ij} + \underbrace{\sum_{v=1}^{N_v} (\gamma_v - \gamma_v^*) z_{v,k(ij)}}_{\mathcal{E}_{k(ij)}} + \mathcal{E}_{k(ij)} \end{aligned} \quad (12)$$

Siden korreksjonen for kovariater bygger på estimater som har usikkerhet, vil det være igjen en rest som inngår i gjennomsnittet og residualvariansen som angitt i ligningen som vist ovenfor. De justerte dataene som kan analyseres for gjennomsnitt og tidstrend vil altså være:

$$\log(C_{ijk}^*) = \mu^* + b \cdot x_i + \varepsilon_i + \alpha_j + \tau_{ij} + \mathcal{E}_{k(ij)}^* \quad (13)$$

Når en skal vurdere utsagnskraft for test av tidstrender vil det mest konservative være å aggregere data til ett gjennomsnitt pr. år og så analysere den resulterende serien via enkel regresjon mot år. Hvis overvåkningsprogrammet består av data fra  $S$  stasjoner og det tas  $R$  replikate prøver fra hver stasjon, vil aggregeringen beskrives ved:

$$y_i = \frac{1}{SR} \sum_{j=1}^S \sum_{k=1}^R \log(C_{ijk}^*) = \underbrace{\mu^* + \frac{1}{S} \sum_{j=1}^S \alpha_j + b \cdot x_i + \varepsilon_i}_{\mu} + \frac{1}{S} \sum_{j=1}^S \tau_{ij} + \underbrace{\frac{1}{SR} \sum_{j=1}^S \sum_{k=1}^R \mathcal{E}_{k(ij)}^*}_{\varepsilon_i} \quad (14)$$

Etter aggregering til en enkel tidsserie har vi altså en modell som kan skrives slik:

$$y_i = \mu' + b \cdot x_i + \varepsilon_i' \quad (15)$$

med residualer  $\varepsilon_i'$  i forhold til tidstrenden. Residualene for gjennomsnitt av replikater pr. år og stasjon i forhold til et tenkt sant årsgjennomsnitt over alle mulige prøvetakingssteder vil ha varians:

$$Var_{si} = \left( \sigma_{si}^2 + \frac{1}{R} \sigma_r^2 \right) \quad (16)$$

Årsgjennomsnittene som middelv verdier over de utvalgte stasjonene i forhold til tidstrenden vil ha varians

$$Var_{\bar{A}_r} = \sigma_t^2 + \frac{1}{S} \left( \sigma_{st}^2 + \frac{1}{R} \sigma_r^2 \right) \quad (17)$$

hvor

$\sigma_t^2$  er varians for fluktuasjoner fra år til år i gjennomsnitt over stasjoner sammenlignet med tidstrend.

$\sigma_{st}^2$  er varians for fluktuasjoner mellom reelle forventningsverdier på forskjellige stasjoner innenfor samme år.

$\sigma_r^2$  er replikat-variens, dvs. varians mellom replikate prøver tatt på samme stasjon og år

Det antas her at prøvene tas med minst ett års mellomrom, slik at  $x_i$  angir hele år, og at alt materiale til  $R$  replikate prøver pr. stasjon og år samles inn fra samme tid og sted. Dersom det er stor systematisk sesongvariasjon, bør prøvene tas på samme tid hvert år for å minimere variansen. Hvis det er stor uregelmessig sesongvariasjon, vil det inngå i varians for interaksjon stasjon\*år, og da kan variasjon mellom analyserte prøver i forskjellige år reduseres ved å fordele innsamlingen av prøvematerialet på flere tidspunkter hvert år. Eventuelt kan innsamlingen også spres på flere lokaliteter i nærheten av stasjonen.

Hvis  $S > 1$  eller  $R > 1$ , kan det å aggregere til en enkel tidsserie være en konservativ test mht. til å finne en tidstrend, fordi antall frihetsgrader i residualestimatet underestimeres. Hvis  $S > 1$  og  $\sigma_{st}^2$  er det helt dominerende leddet, vil de  $T \cdot S$  gjennomsnittene for år\*stasjon i praksis være uavhengige observasjoner, og variansestimater har i realiteten  $T \cdot S \cdot 2$  frihetsgrader. Hvis  $R > 1$  og  $\sigma_r^2$  er det dominerende leddet, vil enkelprøvene utgjøre  $T \cdot S \cdot R$  uavhengige observasjoner, og residualvariansen bestemmes da med  $T \cdot S \cdot R - 2$  frihetsgrader. For en mellomstasjon hvor alle tre bidrag er vesentlige, kan effektivt antall frihetsgrader anslås ut fra estimerte varianskomponenter (Se Vedlegg F). Det krever litt mer kompleks bearbeiding av resultatene, men er innebygd i noen statistiske verktøy, bl.a. i Statistica. I denne omgang antas konservativt at antall frihetsgrader er  $N - 2$ .

Alternativt kan vi anta at gjennomsnittet for hver kombinasjon av stasjon og år kan behandles som uavhengige observasjoner. Hvis data korrigeres for forskjeller mellom stasjoner har vi da

$$y_{ij} = \frac{1}{R} \sum_{k=1}^R \log(C_{ijk}^*) - \alpha_j = \mu^* + b \cdot x_i + \varepsilon_i + \tau_{ij} + \underbrace{\frac{1}{R} \sum_{k=1}^R \varepsilon_{k(ij)}^*}_{\varepsilon_{ij}} \quad (18)$$

Tidstrenden kan da analyseres i en kovariansanalyse og testes mot varians for residualene  $\varepsilon_{ij}$ , som har forventningsverdi:

$$Var_{\bar{A}_r} = \sigma_t^2 + \sigma_{st}^2 + \frac{1}{R} \sigma_r^2 \quad (19)$$

Antagelse om uavhengighet innebærer at  $\sigma_t^2 = 0$ .  $Var_{\bar{A}_r}$  estimeres med omtrent  $(N-1)(S-1)$  frihetsgrader dersom  $\sigma_t^2$  er negligierbar. Hvis det ikke er tilfelle, vil bidraget herfra inngå i  $\varepsilon_{ij}$  som vist og gi korrelerte residualer, og da vil effektivt antall frihetsgrader være lavere.



Hvis innsamlingen av materiale til en prøve fordeles på  $M$  lokalieter og/eller tidspunkter på en slik geografisk avstand eller tidsskala at variansbidraget  $\sigma_{st}^2$  gjelder mellom de  $M$  delprøvene, vil det å blande dem redusere variasjonen mellom de prøvene som analyseres. Vi har da  $S \cdot R$  analyserte prøver, men  $S \cdot R \cdot M$  innsamlinger av data. Utrykket for varians av årsmidlet blir:

$$Var_{\hat{A}_r} = \sigma_t^2 + \frac{1}{S} \left( \frac{1}{M} \sigma_{st}^2 + \frac{1}{R} \sigma_r^2 \right) \quad (20)$$

og for varians mellom avvik for gjennomsnitt fra ulike kombinasjoner av stasjon og år fra jevne trender og stasjonsgjennomsnitt:

$$Var_{Stasjon, \hat{A}_r} = \sigma_t^2 + \frac{1}{M} \sigma_{st}^2 + \frac{1}{R} \sigma_r^2 \quad (21)$$

Fordi avvikene mellom ulike kombinasjoner av stasjon og år også sannsynligvis er skjevfordelt i retning av lognormal fordeling, på samme måte som for individer, vil fysisk blanding av delprøver før analyse være noe mindre effektivt enn å midle logtransformerte verdier fra analyser av hver delprøve. antall lokaliter  $M$  i disse ligningene skulle derfor egentlig være opphøyd i en eksponent  $1/(1 + \sigma_{st}^2/7.7)$ , på samme måte som beskrevet i Vedlegg H. Fordi  $\sigma_{st}^2$  er estimert til ca 0.09 blir eksponenten imidlertid 0.99, slik at det ikke har noen betydning.

Siden analysene av dioksindata indikerte at varianskomponenten  $\sigma_{st}^2$  er liten i forhold til  $\sigma_{st}^2$ , kan det være forsvarlig å basere analysen av utsagnskraft på at hvert stasjons\*år-gjennomsnitt er en uavhengig observasjon og neglisjere  $\sigma_t^2$ , men i første omgang brukes den mest konservative fremgangsmåten.

## Vedlegg C. Hypotesetesting – signifikansnivå og teststyrke

Statistisk hypotese-testing vil si å undersøke om det er grunn til å forkaste en null-hypotese ( $H_0$ ) til fordel for en alternativ hypotese ( $H_a$ ). Det gjøres ved å beregne en testverdi ut fra data, og sammenligne den med en kritisk verdi. Dersom det er liten sannsynlighet for at vi skulle få den beregnede testverdien dersom nullhypotesen er sann, blir den forkastet. Vi ønsker å ha lav sannsynlighet for å ta feil, og opererer med to sannsynlighetsnivåer:

Signifikansnivå  $\alpha$ : Sannsynlighet for å forkaste nullhypotesen feilaktig dersom den er sann.

Teststyrke  $1-\beta$ : Sannsynlighet for å forkaste nullhypotesen dersom den er usann.

Dette kan settes opp som en tabell over fordeling av sannsynlighet for ulike utfall:

Metoden:	Virkeligheten:	
	Hvis $H_0$ er sann	Hvis $H_a$ er sann
Aksepterer nullhypotesen	$1-\alpha$	$\beta$ (feil av type II)
Forkaster nullhypotesen	$\alpha$ (feil av type I)	$1-\beta$

Signifikansnivået  $\alpha$  er den risiko (sannsynlighet) som aksepteres for å ta feil ved å påstå at nullhypotesen er usann dersom den er faktisk er sann. Det innebærer ingen krav til utagskraft, dvs. evne til å skille mellom alternativer. Signifikansnivået kan velges som en bestemt, konstant sannsynlighet, uavhengig av datamengde.

Utsagskraft sikres ved å kreve lav  $\beta$  (høy teststyrke  $1-\beta$ ) samtidig med et bestemt signifikansnivå  $\alpha$ . Den samlede sannsynligheten for å ta feil blir da  $\leq \max(\alpha, \beta)$ . Høy teststyrke kan imidlertid ikke oppfylles generelt for alle alternativer til  $H_0$ . Teststyrken er pr. definisjon sannsynligheten for å forkaste nullhypotesen dersom den er usann, og vil være større jo mer sannheten er forskjellig fra nullhypotesen. Krav til at  $\beta$  skal være under en viss grense må derfor kobles til at forskjellen fra  $H_0$  er over en viss grense.

## Vedlegg D. Teststyrke for t-tester

Ut fra data beregnes en verdi  $\hat{y}$ , f.eks. et gjennomsnitt eller en regresjonskoeffisient, og et usikkerhets-estimat  $s$  for denne verdien. Nullhypotesen  $H_0$  går ut på at  $\hat{y}$  har en bestemt forventningsverdi  $\mu_0$ , og testes ved å se på forholdstallet  $t = (\hat{y} - \mu_0)/s$ . Hvis  $H_0$  er sann, vil  $t$  være t-fordelt med forventnings-verdi 0, med frihetsgrader  $v$  bestemt av mengden data. Hvis  $H_0$  er feilaktig, vil forventningsverdien for  $t$  være forskjellig fra null, og  $t$  er fordelt etter en såkalt desentral  $t$ -fordeling med desentralitets-parameter  $\delta = (\mu - \mu_0)/\sigma$  hvor  $\sigma$  er forventningsverdi for  $s$ . Nullhypotesen er at  $\delta=0$  (tosidig test) eller  $\delta \leq 0$  (ensidig test). Hvis nullhypotesen er sann, skal sannsynligheten for å forkaste den være  $\leq \alpha$ . La  $r(\delta)_{p,v}$  betegne p-fraktil i den desentrale t-fordelingen med  $v$  frihetsgrader, dvs. at det er en sannsynlighet  $p$  for at en tilfeldig verdi som følger fordelingen skal være mindre enn  $r(\delta)_{p,v}$ . Med  $\delta=0$  fås den vanlige sentrale t-fordelingen. Teststyrken for gitt  $\delta \neq 0$  er da gitt som sannsynligheten for at en verdi fra den aktuelle desentrale  $t$ -fordelingen skal overstige den kritiske verdien i den sentrale  $t$ -fordelingen. For å beregne teststyrke for gitt  $\mu - \mu_0$  brukes et estimat for  $\sigma$  i uttrykket for  $\delta$ . Basert på et estimat med stor usikkerhet blir selvsagt også teststyrken usikkert beregnet. Det finnes også et alternativ som er basert bare på den sentrale t-fordelingen. Hypotese-testingen kan oppsummeres slik:

Null-hypotese $H_0$	Forkastes med signifikans-nivå $\alpha$ hvis:	Sammenheng mellom avvik og teststyrke ut fra desentral t-fordeling	Tilnærmet sammenheng mellom avvik og teststyrke basert på sentral t-fordeling:	Alternative hypotese:
$\delta=0$	$ t  \geq t(0)_{1-\alpha/2,v}$	$t(\delta)_{\beta,v} = t(0)_{1-\alpha/2,v}$	$ \delta^*  \geq (t(0)_{1-\beta,v} + t(0)_{1-\alpha/2,v})$	$\delta \neq 0$
$\delta \leq 0$	$t \geq t(0)_{1-\alpha,v}$	$t(\delta)_{\beta,v} = t(0)_{1-\alpha,v}$	$\delta^* \geq (t(0)_{1-\beta,v} + t(0)_{1-\alpha,v})$	$\delta > 0$

Beregning av teststyrke basert på sentral t-fordeling gir en tilnærmet verdi  $\delta^*$  for gitt  $\beta$  og  $\alpha$ . Den tilnærmede sammenhengen er lettere å bruke fordi den sentrale t-fordelingen er mer tilgjengelig i ulike typer beregningsverktøy og i statistiske tabeller. Den gir omtrent identiske resultater med bruk av desentral  $t$ -fordeling for tester med mer enn 30 frihetsgrader, men med få frihetsgrader blir det noe avvik.

En systematisk sammenligning av de to beregningsmåtene for ulike  $\alpha$ ,  $\beta$  og  $v$  viser at for enveis test med  $\alpha=0.05$  og  $\beta$  i området 0.05 til 0.2 blir evnen til å påvise avvik fra nullhypotesen noe overestimert av den tilnærmede beregningen ved  $v=1$ : for  $\beta=0.1$  til 0.15 gir den tilnærmede beregningen ca. 10 % lavere krav til  $\delta$  enn den som baseres på desentral fordeling. For  $v=2$  blir teststyrken litt overestimert for  $\beta$  i området 0.1 til 0.2, men bare med ca. 1 %. For  $v>2$  blir teststyrken underestimert, dvs. at det beregnes noe for høyt krav til  $\delta$  ved gitt  $\beta$ , men innenfor 4 % ved  $\beta=0.05$ , og innenfor 2 % ved  $\beta \geq 0.1$ .

For enveis test med  $\alpha=0.1$  og  $\beta$  i området 0.05 til 0.2 blir evnen til å påvise avvik fra nullhypotesen alltid underestimert, dvs. at den tilnærmede desentralitetsparameteren  $\delta^*$  blir beregnet litt for høyt i forhold til  $\delta$ . For dette tilfelle kan  $\delta$  beregnes innenfor en nøyaktighet på ca. 2 % ut fra den tilnærmet beregnede  $\delta^*$  som funksjon av  $\beta$  og antall frihetsgrader etter følgende formel:

$$\delta = \delta^* \left( 1 + \frac{0.0089}{\beta^{1.3} v^{0.9}} \right)^{-1} \tag{22}$$

Teststyrker som presenteres i denne rapporten er basert på denne korreksjonen, som altså bare er gyldig for enveis test med  $\alpha=0.1$ , for andre testnivåer kan det finnes lignende sammenhenger.

# Vedlegg E.    Utsegnskraft ved hypotesetesting av tidstrend og gjennomsnittsnivåer

## Generelt grunnlag

Utgangspunktet for analysen av utsegnskraft er  $n$  tallpar  $(y_i, x_i)$ , hvor logtransformerte måleverdier  $y$  følger en lineær sammenheng<sup>1</sup> med tidspunkt  $x$  og har normalfordelte avvik fra trenden  $\varepsilon_i$  som er uavhengige og normalfordelte med gjennomsnitt 0 og varians  $\sigma^2$ . I forhold til beskrivelsen i Vedlegg B skriver vi modellen på sentrert form uten markeringer (''):

$$y_i = y_0 + b \cdot (x_i - \bar{x}) + \varepsilon_i; \quad i = 1, \dots, n \quad \text{med} \quad \bar{x} = \sum_{i=1}^n x_i \quad (23)$$

På grunnlag av data estimeres modellparametrene i denne modellen slik:

Modellparameter:	Estimat for verdi:	Varians for estimatet
Forventningsverdi $y_0$ i senter (ved $\bar{x}$ )	$\bar{y} = \sum_{i=1}^n y_i$	$V(\bar{y}) = \frac{\sigma^2}{n}$
Residual-variansen $\sigma^2$	$s^2 = \frac{\sum_{i=1}^n \left( y_i - \bar{y} - \hat{b}x_i \right)^2}{n-2}$	
Regresjonskoeffisient $b$	$\hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$	$V(\hat{b}) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$
Forventningsverdi siste år $y_n$ :	$\hat{y}_n = \bar{y} + b(x_n - \bar{x})$	$V(\hat{y}_n) = \sigma^2 \left( \frac{1}{n} + \frac{(x_n - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)$

## Test på hypoteser om regresjonskoeffisient (se kapittel 9.2.2)

For å teste hypoteser om regresjonskoeffisienten i forhold til en antatt verdi  $b_0$  brukes forholdstallet

$$t = \frac{\hat{b} - b_0}{s} \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (24)$$

Nullhypotese  $H_0: b = b_0$  forkastes hvis  $|t| > t(0)_{(1-\alpha/2), n-2}$  (tosidig test)

Nullhypotesen  $H_0: b \leq b_0$  forkastes hvis  $t > t(0)_{(1-\alpha), n-2}$  (ensidig test)

<sup>1</sup> Analyse av utsegnskraft gjøres for lineær tidstrend (evt. på logskala), men ved fremtidig testing kan det selvsagt være aktuelt også med andre typer modeller.

Her er  $t(0)_{p,n-2}$  øvre p fraktil i sentral  $t$ -fordeling med  $n-2$  frihetsgrader. Denne fordelingen gjelder dersom  $b=b_0$ . Hvis nullhypotesen ikke er sann, er teststørrelsen  $t$  fordelt etter den desentrale  $t$ -fordelingen med  $n-2$  frihetsgrader. Denne fordelingen er bestemt av en desentralitetsparameter, som er forholdet mellom forskjell mellom virkelig verdi og hypoteseverdi og usikkerheten for estimatet for virkelig verdi:

$$\delta = \frac{\beta - \beta_0}{\sqrt{V(\hat{\beta})}} \quad (25)$$

Mens den sentrale  $t$ -fordelingen har sin topp ved  $t=0$ , har den desentrale fordelingen sin topp omkring  $t=\delta$ , slik at det er mindre sannsynlighet for verdier omkring 0 jo mer  $\delta$  avviker i positiv eller negativ retning. Teststyrken for gitt regresjonskoeffisient  $b$  øker med økende størrelse på desentralitetsparameteren og med økende antall frihetsgrader.<sup>1</sup>

Hvis tidsserien har én verdi hvert år over et tidsrom på  $T$  år, dvs. med  $n=T$  observasjoner hvor  $x_i=i$  for  $i=1, \dots, n$ , har vi:

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n i^2 - \frac{1}{n} \left( \sum_{i=1}^n i \right)^2 \\ &= \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4} \\ &= \frac{(n-1)n(n+1)}{12} \end{aligned} \quad (26)$$

Hvis observasjonene ikke er gjort hvert år, men med periode  $h$ , og vi har en enkelt serie av  $n$  verdier over et tidsrom  $T = (n-1)h + 1$  år, slik at  $x_i = (i-1) \cdot h + 1$  for  $i=1, \dots, n$ , har vi:

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n h^2 (i - \bar{i})^2 = h^2 \frac{(n-1)n(n+1)}{12} \quad (27)$$

Hvis serien består av årsgjennomsnitt over  $S$  stasjoner, er  $\sigma^2 = Var_{\bar{X}_r}$  som beskrevet i Vedlegg B ligning (17), og estimatet for  $\sigma^2$  har  $v=n-2$  frihetsgrader. Hvis vi innfører  $n_r =$  antall tidspunkter som i dette tilfelle er det samme som antall verdier  $n$ , kan desentralitetsparameteren da skrives:

$$\delta = (b - b_0)h \frac{S(n_r - 1)n_r(n_r + 1)}{\sqrt{12 \left( S\sigma_i^2 + \sigma_{st}^2 + \frac{1}{R}\sigma_r^2 \right)}} \quad (28)$$

<sup>1</sup> Teststyrken for  $t$ -test som funksjon av  $\gamma$ , antall frihetsgrader  $v$  og valgt signifikansnivå  $\alpha$  vil tilsvare teststyrken for en enkel  $t$ -test basert på  $v+1$  observasjoner med standardisert effekt, dvs. forhold mellom avvik i gjennomsnitt og standardavvik for enkeltverdi på  $\gamma/\sqrt{v+1}$ .

Hvis det antas at  $\sigma_i^2 = 0$ , slik at data kan analyseres som  $S$  uavhengige observasjoner for hvert tidspunkt, blir  $\sigma^2 = Var_{st}$  som beskrevet i Vedlegg B, ligning (16)<sup>1</sup> og antall verdier  $n = S \cdot n_t$  hvor  $n_t$  = antall tidspunkter. Det gir:

$$\sum_{i=1}^n (x_i - \bar{x})^2 = S \sum_{k=1}^{n_t} h^2 (k - \bar{k})^2 = S \cdot h^2 \frac{(n_t - 1)n_t(n_t + 1)}{12} \quad (29)$$

Data justeres da for permanente stasjonsforskjeller i en kovariansanalyse, slik at antall frihetsgrader er  $\nu = S \cdot n_t - (S - 1) - 2 = (n_t - 1)S - 1$  og desentralitetsparameteren blir

$$\delta = (b - b_0)h \sqrt{\frac{S(n_t - 1)n_t(n_t + 1)}{12 \left( \sigma_{st}^2 + \frac{1}{R} \sigma_r^2 \right)}} \quad (30)$$

Forholdet  $b/\sigma$  har dimension  $tid^1$  og  $h$  har dimension  $tid$ , slik at  $\delta$  er et dimensjonsløst tall. Teststyrken er en funksjon av  $\delta$  og antall frihetsgrader  $\nu$ . Desentralitetsparameteren  $\delta$  er en funksjon av avvik i regresjonskoeffisient  $(b - b_0)$ , overvåkningsprogrammets omfang i tid ( $h$  og  $n_t$ ) og omfanget av prøvetaking ved hvert tidspunkt samt de estimerte varianskomponentene som beskrevet i Vedlegg B. Hvis en har et estimat  $s$  for  $\sigma$  basert på data som beskrevet ovenfor, er det vanlig å sette det inn i uttrykket for desentralitetsparameteren og beregne teststyrken ut fra dette. Bruk av desentral fordeling gir eksakt beregning av teststyrke for gitt  $b - b_0$  hvis antall frihetsgrader er gitt og  $\sigma$  er kjent, men basert på et estimat med usikkerhet blir selvsagt også teststyrken usikkert beregnet. En tilnærmet metode er beskrevet i Vedlegg D.

### Test på nå-tilstand i forhold til klassifiseringsgrense for konsentrasjon (se kapittel 9.2.3)

Hensikten er å klassifisere et geografisk område ved å sammenligne estimat for gjennomsnittskonsentrasjon  $\mu$  med en klassifiseringsgrense  $\mu_c$ , for eksempel for å revurdere kostholdsråd. Fordi det er hensiktsmessig å gjøre analysen på logtransformerte data, antas at klassifiseringsgrensen gjelder geometrisk middelværdi. Nullhypotesen  $H_0$  er at midlere konsentrasjon  $\mu$  overskrider klassifiseringsgrensen  $\mu_c$ . Hvis grensen ikke er overskredet, men det faktiske gjennomsnittet er  $\mu = f \cdot \mu_c$  hvor  $f < 1$ , ønsker vi at  $H_0$  skal forkastes og området klassifiseres som å være innenfor grensen med så stor sannsynlighet som mulig. På logskala blir det  $\log(\mu) = \log(f) + \log(\mu_c)$ .

Hvis det er lite reduksjon pr. år, er det rimelig å klassifisere etter gjennomsnitt  $\bar{y}$  for logtransformerte konsentrasjoner som estimat for log av geometrisk middelværdi  $y_m = \log(\mu)$ . For en enkel dataserie, beregnet som årsgjennomsnitt over  $S$  stasjoner, er variansen for  $\bar{y}$  gitt ved:

$$V(\bar{y}) = \frac{\sigma_t^2 + \frac{1}{S} \left( \sigma_{st}^2 + \frac{1}{R} \sigma_r^2 \right)}{n_t} \quad (31)$$

<sup>1</sup> I ligning (16) er estimert varians for fluktasjoner mellom år som er felles over stasjoner ikke tatt med. En kunne også brukt ligning (19), som åpner for å betrakte estimert ikke-signifikant varians mellom år som et tilleggskorrigering til uavhengig variasjon mellom stasjoner. Det gir noe mer konservativt resultat når det gjelder evnen til å påvise endringer eller klassifisere.

og hvis  $\sigma_t^2 = 0$  slik at data kan analyseres som  $S$  uavhengige observasjoner pr. år:

$$V(\bar{y}) = \frac{\sigma_{st}^2 + \frac{1}{R}\sigma_r^2}{Sn_t} \quad (32)$$

Hvis det er en tydelig nedgang i løpet av overvåkingsperioden, kan det være aktuelt å klassifisere ved å sammenligne estimert forventningsverdi av logtransformerte konsentrasjoner  $\hat{y}_n$  for siste tidspunkt med klassifiseringsgrensen. For ekvidistante  $x$  er variansen for denne forventningsverdien gitt bare av antall data, uavhengig av  $T$  og  $h$  for gitt  $n$ .

For en enkel tidsserie med  $n_t = n$  tidspunkter finner vi ved å sette  $\sum_{i=1}^n (x_i - \bar{x})^2 = \frac{(n-1)n(n+1)}{12}$  og  $\bar{x} = (n+1)/2$  inn i uttrykket for  $V(\hat{y}_n)$ :

$$V(\hat{y}_n) = \frac{\sigma_t^2 + \frac{1}{S}\left(\sigma_{st}^2 + \frac{1}{R}\sigma_r^2\right)}{n_t} \left(1 + 3\frac{(n_t-1)}{(n_t+1)}\right) \quad (33)$$

Hvis  $\sigma_t^2 = 0$ , slik at det er  $S$  uavhengige observasjoner pr. tidspunkt ( $n = S \cdot n_t$  hvor  $n_t =$  antall tidspunkter) blir uttrykket for variansen i stedet:

$$V(y_n) = \frac{\sigma_{st}^2 + \frac{1}{R}\sigma_r^2}{Sn_t} \left(1 + 3\frac{(n_t-1)}{S(n_t+1)}\right) \quad (34)$$

Nullhypotesen  $\mu \geq y_C$  ("klassifiseringsgrensen overskrides") blir forkastet hvis

$$t = \frac{y_C - \tilde{y}}{\sqrt{V(\tilde{y})}} > t(0)_{(1-\alpha), \nu} \quad (35)$$

hvor er  $\tilde{y}$  er estimat for gjennomsnittet  $y_m$ , dvs. henholdsvis  $\bar{y}$  eller  $\hat{y}_n$ . På samme måte som for påvisning av tidstrend bygger vurdering av teststyrke på hvordan testverdien  $t$  er fordelt ut fra en desentralitetsparameter  $\delta$ :

$$\text{Liten tidstrend, klassifisering ut fra gjennomsnitt} \quad \delta = \frac{y_C - y_m}{\sqrt{Var(\bar{y})}}$$

$$\text{Tydelig tidstrend- klassifisering ut fra slutten av tidsserien} \quad \delta = \frac{y_C - y_m}{\sqrt{Var(\hat{y}_n)}}$$

### Virkning av korrelerte residualer

Beskrivelsen ovenfor gjelder enkel regresjonsanalyse eller kovariansanalyse på data med uavhengige residualer (=avvik fra modellen). Hvis grunnlagsdata har delvis korrelerte residualer, forutsettes da at data aggregeres inntil en får uavhengige residualer, f.eks. ved å analysere på åsgjennomsnitt på tvers av stasjoner med korrelerte residualer, eller på gjennomsnitt av observasjoner som er så nær i tid at de har korrelerte residualer.

En annen måte å håndtere korrelerte residualer er å bruke variansanalyse for å skille ulike varianskomponenter fra hverandre, og estimere det effektive antall frihetsgrader ut fra resultatet, eller

estimere korrelasjon mellom observasjoner som ligger nær i tid og redusere antall frihetsgrader ut fra dette. Dersom residualene er korrelert og dette blir neglisjert, vil virkningen først og fremst være å overdrive effektivt antall frihetsgrader. I tillegg vil estimatet  $s$  ha tendens til å bli for lav fordi det beregnes på et for stort antall frihetsgrader, men det er uvesentlig. Estimatet for  $b$  vil ikke få endret forventingsverdi.



## Vedlegg F. Beregning av varianskomponenter fra variansanalyser

Varians- og kovariansanalyser produserer såkalte ANOVA-tabeller, hvor varians for de ulike effektene sammenlignes med sannsynlig virkning av tilfældige variasjoner på mindre skala for å se om det er tegn til reelle effekter. I slike tabeller inngår estimater for midlere kvadratavvik (*Mean Square* eller MS) for de forskjellige effektene. MS er et uttrykk for den totale variansen for effekten, skalert slik at den skal bli lik residual-variansen for enkeltobservasjon hvis det ikke er noen effekt, og all variasjon bare er restvirkinger av tilfeldig variasjon mellom enkeltprøver. For modellen som er beskrevet i Vedlegg B, og med et fullstendig og balansert datasett, vil forventet verdi for de forskjellige *Mean Squares* som funksjon av de forskjellige variansbidragene til enkeltprøver være gitt ved uttrykkene nedenfor<sup>1</sup>, og estimatene vil da være uavhengige stokastiske variable.

	Uttrykk for <i>mean square</i> (MS) i ANOVA-tabell	frihetsgrader
Fluktuasjon i årsmidler	$EMS_{\hat{A}_r} = SR\sigma_i^2 + R\sigma_{st}^2 + \sigma_r^2$	T-1
Uavhengige fluktuasjoner for år*stasjon (interaksjon)	$EMS_{\hat{A}_r, stasjon} = R\sigma_{st}^2 + \sigma_r^2$	T(S-1)
Variasjon mellom replikater	$EMS_r = \sigma_r^2$	TS(R-1)

Dersom datasettet ikke er komplett (mangler data for noen kombinasjoner) eller er ubalansert (varierende antall replikater) blir koeffisientene annerledes, og de forskjellige estimatene vil ikke alltid være uavhengige, men analysen gir en tilsvarende tabell, slik at vi alltid har:

	Uttrykk for <i>mean square</i> (MS) i ANOVA-tabell	frihetsgrader
Fluktuasjon i årsmidler	$EMS_{\hat{A}_r} = k_1\sigma_i^2 + k_2\sigma_{st}^2 + \sigma_r^2$	$\nu_{\hat{A}_r}$
Uavhengige fluktuasjoner for år*stasjon (interaksjon)	$EMS_{\hat{A}_r, stasjon} = k_3\sigma_{st}^2 + \sigma_r^2$	$\nu_{\hat{A}_r, stasjon}$
Variasjon mellom replikater	$EMS_r = \sigma_r^2$	$\nu_r$

<sup>1</sup> Eventuelle konstante forskjeller mellom stasjoner forandrer ikke på dette.

Gitt estimator for de forskjellige *Mean Squares* (MS) kan vi da beregne variansbidragene for år og år\*stasjon slik:

	Uttrykk for <i>mean square</i> (MS) i ANOVA-tabell
Fluktuasjon i årsmidler	$\sigma_i^2 = \frac{MS_{\hat{A}r} - \frac{k_2}{k_3} MS_{\hat{A}r, stasjon} - \left(1 - \frac{k_2}{k_3}\right) MS_r}{k_1}$
Uavhengige fluktuasjoner for år*stasjon (interaksjon)	$\sigma_{st}^2 = \frac{MS_{\hat{A}r, stasjon} - MS_r}{k_3}$

En variansanalyse gir estimator for varians for de forskjellige tilfeldige effektene som inngår i modellen, det vil si estimator for 'expected mean squares' eller EMS. Dersom de tilfeldige utslagene er normalfordelt, vil disse estimatorne vil være  $\chi^2$ -fordelte stokastiske variable med frihetsgrader bestemt av hvor mange uavhengige utslag som er kombinert i estimatorne. Hvis en har en serie uavhengige slike variansesimulator  $s_i^2$  for  $i=1, \dots, N$  med frihetsgrader  $v_i$  og beregner en lineær kombinasjon av estimatorne:

$$S^2 = \sum_{i=1}^N k_i s_i^2$$

kan  $S^2$  betraktes tilnærmet som en  $\chi^2$ -fordelt stokastisk variabel med antall frihetsgrader gitt ved uttrykket:

$$v_s = \frac{S^4}{\sum_{i=1}^N \left(k_i s_i^2\right)^2 v_i^{-1}}$$

Dette kalles Satterthwaites tilnærmelse. Hvis alle  $k_i$  er positive vil  $v_s$  bli større enn alle  $v_i$ . Hvis  $k_i$  har forskjellig fortegn, slik at  $S^2$  delvis er en forskjell mellom variansesimulator, kan antall frihetsgrader bli lavere, og i noen tilfelle  $<1$ .

## Vedlegg G.    Presisjon for variansestimater som funksjon av antall frihetsgrader

Hvis  $s^2$  er et variansestimat basert på  $n$  normalfordelte uavhengige verdier med varians  $\sigma^2$ , er forholdet  $(n-1)s^2/\sigma^2$  fordelt etter en  $\chi^2$ -fordeling med  $v=n-1$  frihetsgrader og vil med sannsynlighet  $1-p$  ligge innenfor grensene

$$\chi^2_{p/2,v} < v \cdot s^2 / \sigma^2 \leq \chi^2_{1-p/2,v} \tag{36}$$

hvor  $\chi^2_{\alpha,v}$  er  $\alpha$ -fraktilen i den kumulative  $\chi^2$ -fordelingen. Konfidensgrensene for  $\sigma$  finnes ved å snu denne relasjonen, dvs.:

$$\frac{v \cdot s^2}{\chi^2_{1-p/2,v}} \leq \sigma^2 \leq \frac{v \cdot s^2}{\chi^2_{p/2,v}} \tag{37}$$

En slik fordeling med få frihetsgrader vil ha ganske stor spredning, spesielt i retning av lave verdier. Med 4 frihetsgrader vil 80 % konfidensintervall for den reelle variansen  $\sigma^2$  gå fra  $0.52 \cdot s^2$  til  $3.76 \cdot s^2$  og medianen for forholdet  $s^2/\sigma^2$  er 0.84, dvs. at reell varians i 50 % av tilfellene vil være  $>1.19 \cdot s^2$ . Det er derfor stor sannsynlighet for at estimert varians blir mindre enn den reelle, og at behov for data blir undervurdert, dvs. at teststyrken blir lavere enn beregnet.

**Tabell 34** viser 10%, 50 % og 90 % fraktiler for forholdet mellom reell varians og estimert varians som funksjon av antall frihetsgrader. Det fremgår at ved mindre enn 5 til 10 frihetsgrader er det overveiende sannsynlig at reell varianser mer enn 20 % større enn den estimerte. Hvis variansestimatet har 3 frihetsgrader, vil den reelle variansen i 10 % av tilfellene være mer enn 5 ganger større enn den estimerte. Det er først omkring 100 frihetsgrader at variansestimatet begynner å bli noenlunde presist.

Dette må tas i betraktning når en baserer statistisk design på usikre variansestimater. Dersom reell varians er 2 ganger større enn estimert, vil behovet for data være omtrent dobbelt så stort som estimert for å oppnå en definert teststyrke for en bestemt effekt, og ved gitt datamengde vil den estimerte teststyrken oppnås først ved en effekt som er 40 % større enn den som er brukt ved beregning av teststyrken.

**Tabell 34.** 80 % konfidensintervaller og median for varians i forhold til variansestimat som funksjon av antall frihetsgrader.

Antall frihetsgrader	90 % nedre konfidensgrense	Median	90 % øvre konfidensgrense
1	0.37	2.20	63.33
2	0.43	1.44	9.49
3	0.48	1.27	5.13
4	0.51	1.19	3.76
5	0.54	1.15	3.10
10	0.63	1.07	2.06
20	0.70	1.03	1.61
30	0.75	1.02	1.46
40	0.77	1.02	1.38
50	0.79	1.01	1.33
100	0.84	1.01	1.21
200	0.88	1.00	1.14

## Vedlegg H. Statistiske egenskaper for blandprøver når konsentrasjoner er log-normalt fordelt.

Ved analyse av miljøgifter i biota er det vanlig å analysere blandprøver av flere individer for å få et bedre bilde av gjennomsnittskonsentrasjonen til gitt analysekosnad. Enkeltest individer kan ha ganske stor spredning i verdier, og ved å blande materiale fra flere individer vil avvikene jevnes ut.

Hvis konsentrasjonene på individnivå er symmetrisk fordelt med varians  $\sigma_i^2$ , vil blandprøver av  $N$  tilfeldig valgte individer ha varians  $\sigma_i^2/N$  rundt gjennomsnittet pga. utvalgsusikkerheten. Med en slik fordeling vil median være lik aritmetisk gjennomsnitt, og ha lik verdi for individer og blandprøver.

Inndeltid er det vanligvis stor spredning i konsentrasjon mellom individer, slik at det forekommer høye konsentrasjoner som er flere ganger større enn medianen mens variasjonen i motsatt retning er begrenset av at konsentrasjonene er  $\geq 0$ , og konsentrasjonene vil ha skjev fordeling. I praksis er da gjerne logtransformerte konsentrasjoner mer symmetrisk fordelt, og beskrives ofte godt av en normalfordeling, dvs. at konsentrasjonene er lognormalt fordelt.

Konsentrasjonen i en blandprøve vil være et aritmetisk gjennomsnitt av individuelle konsentrasjoner, og det vil da ofte være noen få høye verdier som dominerer. I en slik situasjon vil blandprøver med lik mengde fra hvert individ ha samme gjennomsnitt som beregnet gjennomsnitt av analyser på hvert av de samme individene, men medianen vil være lavere enn gjennomsnittet, og ikke likt for analyse av individuelle prøver enkeltvis og tilsvarende blandprøver.

Tilfeldig variasjon i konsentrasjoner i blandprøver pga. utvalgsusikkerhet vil ikke avta like effektivt med økende antall individer som ved en symmetrisk fordeling. Også blandprøvene vil ha skjev fordeling, om enn i mindre grad enn de individuelle prøvene, slik at logtransformasjon er aktuelt for å gi mer symmetrisk fordeling. Selv om konsentrasjonen i blandprøvene ikke vil følge noen enkel veldefinert fordeling, kan lognormal fordeling være en god beskrivelse også for dem, i alle fall for å vurdere statistiske mål som bygger på mange blandprøver, f.eks. regresjon mot tid i logtransformerte verdier. For å vurdere hva en oppnår ved blandprøver av ulik størrelse er det nødvendig å vite hvordan variansen avhenger av prøvestørrelsen, dvs. i hvilken grad de individuelle variasjonene jevner seg ut.

For å kartlegge dette er det gjort en numerisk simulering. For ulike kombinasjon av antall individer i prøven ( $N$ ) og varians av  $\log(C)$  for individer ( $\sigma_i^2$ ) er det trukket 10 000 sett av  $N$  enkeltverdier fra lognormal fordeling med gjennomsnitt 0 for  $\log(C)$ , dvs. at medianverdi for konsentrasjon er 1. For hvert sett beregnes aritmetisk gjennomsnitt av konsentrasjonene ( $C$ ) tilsvarende analyse av blandprøve. På basis av de 10 000 blandprøve-konsentrasjonene beregnes så gjennomsnitt på lineær skala, samt gjennomsnitt  $\mu_b$  og varians  $\sigma_b^2$  for logtransformerte blandprøveverdier ved hver kombinasjon av  $N$  og  $\sigma_i^2$ .

Ved ikke-lineær tilpasning av funksjoner til disse resultatene er det funnet at sammenhengen mellom varianser av logtransformerte konsentrasjoner kan uttrykkes ganske presist av:

$$\sigma_b^2 = \frac{\sigma_i^2}{N^{1/(1+\sigma_i^2/7.7)}} \quad (38)$$

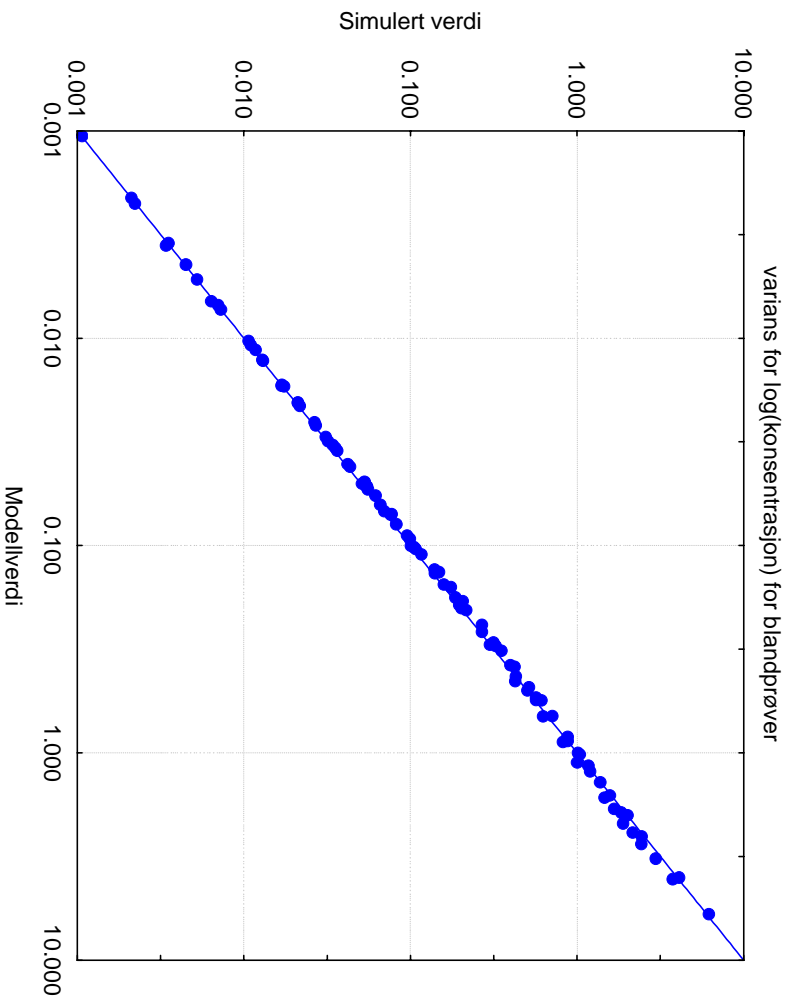
Denne formelen kan brukes direkte til å beregne varians for blandprøver hvis en kjenner varians for enkeltpørve. Hvis en i stedet har estimator for  $\sigma_b^2$  ved gitt  $N$ , kan den brukes iterativt for å finne  $\sigma_i^2$  og deretter også  $\sigma_b^2$  for andre verdier av  $N$ .

I forhold til vanlig utjevning for symmetrisk fordeling kan det defineres et 'effektivt' antall individer ut fra forholdet mellom varianser på logskala:

$$N_{eff} = \frac{\sigma_i^2}{\sigma_b^2} = N^{1/(1+\sigma_i^2/7.7)} \quad (39)$$

Hvis en f.eks. har blandprøver av 50 individer med residualvarians 0.1 på logskala for prøvene, tilsvarer det varians ca. 2.0 på logskala for individnivå (dvs. at ca. 70 % av individene ligger innenfor et intervall fra 0.15 til 7 ganger medianen), og effektivt antall individer ca. 20. Hvis residualvariansen for blandprøver er 0.03 tilsvarer det varians ca. 1.0 på logskala mellom individer, dvs. at 70 % av individene varierer fra 0.37 til 2.7 ganger median.

Figuren nedenfor viser overensstemmelse med simulerte verdier og verdier beregnet av den tilpassede sammenhengen:



Ovenfor er det antatt at det er like store mengder vev fra hvert individ. Hvis også mengden vev varierer mellom individer, vil også variasjonen i mengde inngå i betraktningen. Hva som blir resultatet, vil da avhenge av om variasjon i mengde og konsentrasjon mellom individer i prøven er korrelert med hverandre.

## Vedlegg I. Definisjon av statistiske begreper

Ordnet alfabetisk, i forklaringene er ord som er definert i ellers i listen vist *i kursiv*.

Kovariansanalyse	Utvidet versjon av <i>variansanalyse</i> hvor en sammen med kategoriske faktorer også har med forklaringsvariable som varierer kontinuerlig, beskrevet ved regresjon.
Nullhypotese	Hypotese om statistisk fordeling av observasjoner av en stokastisk variabel, vanligvis om gjennomsnittsnivå for av alle observasjoner eller grupper av observasjoner. Statistisk testing går ut på å se om datamaterialet gir grunnlag for å forkaste nullhypotesen til fordel for en alternativ hypotese. Hvis settet av observasjoner avviker tilstrekkelig fra det som er sannsynlig resultat dersom nullhypotesen er sann, blir nullhypotesen forkastet. Hvis avviket ligger innenfor det som er rimelig å vente som tilfeldig avvik, aksepteres den.
Replikater	Prøver som er tatt fra samme "minsteenh" i prøvetakingssopplegget, her fra samme kombinasjon av tid og sted.
Residualer	Her brukt om forskjellen mellom enkeltobservasjoner og estimert verdi ut fra <i>varians/kovarians</i> -modeller for konstante geografiske forskjeller og jevne tidstrender.
Residualvarians	Her brukt om <i>variansen</i> for <i>residualene</i> i en <i>varians/kovarians</i> -analyse
Signifikansnivå	Sannsynligheten for å forkaste <i>nullhypotesen</i> selv om den er sann. Velges vanligvis å priori: jo lavere signifikansnivået settes, jo mer må observasjonene avvike fra det som kan forventes ut fra nullhypotesen for å forkaste den.
Standardavvik	Kvadratrot av <i>varians</i> , brukt som mål på et typisk avvik fra gjennomsnitt for enkeltobservasjoner. For observasjoner fra en normalfordeling vil 67 % av observasjonene ligge innenfor $\pm$ ett standardavvik fra gjennomsnittet i fordelingen.
Standardfeil	Kvadratrot av <i>varians</i> for et statistisk estimat, for eksempel et gjennomsnitt over flere observasjoner. Middelverdien av $N$ uavhengige observasjoner som har standardavvik $\sigma$ vil ha standardfeil $\sigma/N$ . Standardfeilen er et mål på typisk avvik fra sant gjennomsnitt for estimatet.
Stokastisk variabel	Måleverdi eller observasjon som er influert av 'tilfeldig' eller uforutsigbar variasjon.
Teststyrke ( <i>Power</i> )	Sannsynligheten for å forkaste <i>nullhypotesen</i> dersom den er usann og det er en alternativ hypotese som er sann, for eksempel ved at gjennomsnitt avviker fra det som er antatt i nullhypotesen. Teststyrken er lien dersom dette avviket er lite, og øker med størrelsen på avviket. Teststyrken for et gitt avvik vil øke med antall observasjoner, og minske for et gitt sett av observasjoner når kravet til signifikansnivået skjerpes, dvs. når signifikansnivået velges mindre.
Utsagnskraft	Her oppfattet som datamaterialets evne til å påvise om <i>nullhypotesen</i> er sann eller ikke, dvs. til å gi en høy <i>teststyrke</i> for de avvikene fra nullhypotesen som det er interessante å oppdage, samtidig som <i>signifikansnivået</i> velges på et rimelig lavt nivå.

Varians	<p>Midlere kvadrat av avviket fra forventningsverdi (gjennomsnitt) for en <i>stokastisk variabel</i>. Hvis variabelen har sannsynlighetsfordeling <math>f(x)</math>, dvs. at sannsynlighet for <math>x &lt; x_c</math> er <math>P = \int_{x_{\min}}^{x_c} f(\xi) d\xi</math> vil variansen være gitt ved:</p> <p>Gitt <math>N</math> observasjoner <math>x_i, i=1, \dots, N</math> fra en sannsynlighetsfordeling <math>f(x)</math>,</p>
Variansanalyse	<p>Brukes på datasett hvor observerte verdier er antatt å variere med en eller flere <i>kategoriske faktorer</i> (dvs. med inndeling i et antall grupper). Variansanalysen beskriver data som en sum av ledd som beskriver virkningen av de ulike faktorene alene eller i interaksjon og ulike bidrag til tilfeldig variasjon og estimerer systematiske forskjeller mellom grupper og varians for tilfeldige, uregelmessige variasjoner rundt gruppegjennomsnittene. Se også <i>kovariansanalyse</i></p>